EXPLORING KEY ORIENTATIONS OF SMALL MOLECULES TO DISRUPT PROTEIN-PROTEIN INTERACTIONS

A Dissertation

by

EUNHWA KO

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2012

Major Subject: Chemistry

EXPLORING KEY ORIENTATIONS OF SMALL MOLECULES TO DISRUPT PROTEIN-PROTEIN INTERACTIONS

A Dissertation

by

EUNHWA KO

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Kevin Burgess
Committee Members,	Coran M. H. Watanabe
	Jiong Yang
	Rajesh C. Miranda
Head of Department,	David H. Russell

May 2012

Major Subject: Chemistry

ABSTRACT

Exploring Key Orientations of Small Molecules to Disrupt Protein-protein Interactions.

(May 2012)

Eunhwa Ko, B.S., Chungnam National University, the Republic of Korea;

M.S., Seoul National University, the Republic of Korea

Chair of Advisory Committee: Dr. Kevin Burgess

Protein-protein interactions (PPIs) are attractive targets because of their therapeutic potential. One approach to design small molecules that can disrupt the PPIs is to use structural information of proteins. With this approach, triazole-based peptidomimetics that mimic β-turn hot-spot regions in neurotrophins were synthesized. The monovalent mimics were assembled into bivalent mimics via a combinatorial method. Three different bivalent mimics were prepared for different studies. Bivalent mimics with long-linkers bound to TrkA or TrkC receptor and showed partial antagonism for the receptors. Other mimics were conjugated with cytotoxic compounds and they were used for TrkC targeted drug delivery. The last group of bivalent mimics previously showed targeted delivery effects for pancreatic cancer cells. In this study, we synthesized Eu-chelated bivalent mimics to perform a competitive binding assay for pancreatic cancer cells.

Previous research in our group focused on design of secondary structures' mimics on rigid scaffolds as "minimalist mimics". We sought to establish structural design criteria for the minimalist mimics, and we wanted to propose that sets of such compounds could mimic local pairs of amino acids in any secondary structures as "universal peptidomimetics". Thus, we designed five compounds, such as oxazoline-, pyrrole-, dyine- "kinked" and "linear" bistrizole-based peptidomimetics, and performed molecular modelings, DFT calculations, and QMD for them to validate our hypothesis.

On the concepts of "minimalist mimics" and "universal peptidomimetics", we developed the Cα – Cβ vector matching program to evaluate preferred orientations of Cα - Cβ coordinates for secondary structures. We applied the program to omegatides and pyrrolinone-pyrrolidine oligomers. The compounds matched better with strands than for helices.

We expanded the Cα – Cβ vector matching idea to a method that ranks preferred conformations of small molecules on any combination of three interface side-chains in all structurally characterized PPIs. We developed a PDB mining program (explores key orientation, EKO) to do this, and EKO applied to pyrrolinone-pyrrolidine oligomers to find targets. EKO found several interesting targets, such as AICAR Tfase, GAPDH, and HIV-1 protease. HIV-1 dimerization inhibition and Zhang-Poorman kinetic assays were performed to validate our hypothesis, and the results showed that pyrrolinone-pyrrolidine derivatives inhibited HIV-1 dimerization.

DEDICATION

To my family

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor Dr Kevin Burgess for his enthusiasm and guidance during my research in his group. I also thank my committee members, Dr Corna Watanabe, Dr Jion Yans, and Dr Rajesh Miranda for their guidance.

I would like to acknowledge Dr Lisa Perez and Dr Thomas Ioeger for their teaching and support for all computational works, and Dr Arora Loudet and Dr Robert Burghardt for their teaching for all cell-based works.

I really want to thank all my colleagues, especially former group members, Drs Amber Schaefer, Yu Li, Jing Liu, David Chen, Andrey Malakhov, Arjun Raghuraman, Dmaytro Fedoseyenko, Liangxing Wu, Cliferson Thivierge, Jiney Jose, Yuichiro Ueno, Junyan Han, Juan Rodriguez-Poirier and Ye Zhu for their friendship and good advice, as well as all the current members for their companionship, especially Sakunchai Khumsubdee.

I would like to thank to Jill Powers for her assistance through all years.

I also thank to Sungwook Cho, Gil Ma, and Changsuk Lee for their support and guidance, and Youngmin Kim, Soyoun Hwang and Hyejin Chun for their friendship.

Finally, I would like to pay special thanks to my family for their love, support and patience.

TABLE OF CONTENTS

Page

Page

LIST OF FIGURES

Page

Page

Page

Page

Page

Page

LIST OF TABLES

Page

Page

LIST OF SCHEMES

CHAPTER I

INTRODUCTION*

1.1 Design of Small Molecules to Resemble Protein Fragments

Proteins and peptides play crucial roles in biology, so they are studied as targets in medicine.[1,2] However, there are several limitations to using them as therapeutic agents, including vulnerability to proteolysis and poor cell permeation. To overcome these problems, chemists have developed molecules that can mimic of peptides with less of these limitations. These "peptidomimetics"[3] have at least some peptidic parts substituted with non-peptide structures.

Peptide analogs can have unnatural fragments that mimic peptidic main-chains, side-chains or both. To mimic only main-chains in peptides, a peptide backbone can be replaced by an ester or thioester (blue circle in Figure 1.1). However, these compounds are uninteresting because they do not have the side-chains that are involved in intermolecular interactions. These structures only form limited *H*-bonding interaction on the backbones.

This dissertation follows the style of the *Journal of American Chemical Society*.

*Reprinted in part with permission from "Minimalist and universal peptidomimetics", Eunhwa Ko, Jing Liu and Kevin Burgess, *Chem. Soc. Rev.,* **2011**, *40*, 4411-4421.

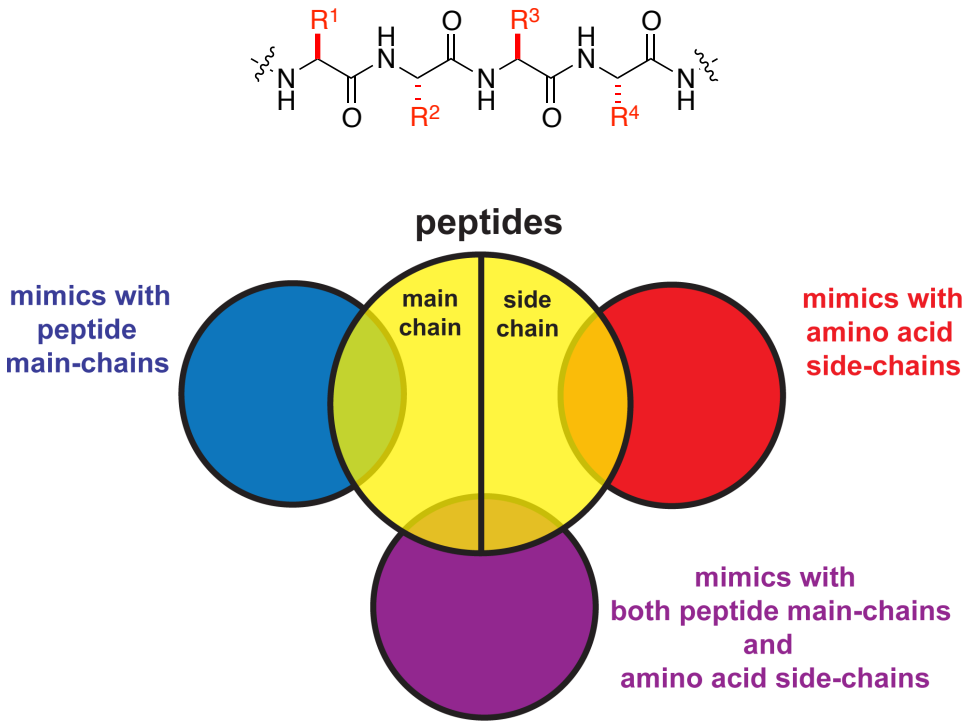


Figure 1.1. Relationships between peptides and peptidomimetics.

Peptide analogs can have unnatural fragments that mimic peptidic main-chains, side-chains or both. To mimic only main-chains in peptides, a peptide backbone can be replaced by an ester or thioester (blue circle in Figure 1.1). However, these compounds are uninteresting because they do not have the side-chains that are involved in intermolecular interactions. These structures only form limited *H*-bonding interaction on the backbones.

Peptidomimetics that mimic both peptidic main-chains and side-chains are more common (purple circle in Figure 1.1). The general strategy is to replace the amide bonds with isosteric groups to decrease the limitations described above (Figure 1.2a). This class of compounds is useful in the design of protease inhibitors, but they are conformationally flexible, and this decreases their binding selectivities and affinities.[4] One method to overcome this problem is to constrain peptide segments. Arora's group applied a main-chain hydrogen bond surrogate (HBS) strategy to design helical

secondary structure mimics as illustrated in Figure 1.2b. The HBS α-helix mimics have peptide backbones constrained via ring-closing metathesis to give fragments that resemble an *N*-terminal intramolecular hydrogen bond between the peptide *i* and *i* + 4 residues.[5] Aube's group designed β-turn mimics that are formed by dipeptides constrained by substituted Aca (6-aminocaproic acid) linkers (Figure 1.2b).[6] In this case, the non-peptidic part is used to lock the overall conformation of the molecule. However, potential applications of these types of compounds as drugs are limited because their large size and high polarities mean that they do not tend to permeate into cells, hence they are inappropriate for intracellular targets. For this reason, the third type of peptidomimetic is intriguing: compounds that mimic only side-chains in analogs (red circle in Figure 1.1) and do not contain {polar} amide functionalities. In our group, we refer to such compounds as minimalist mimics.

a

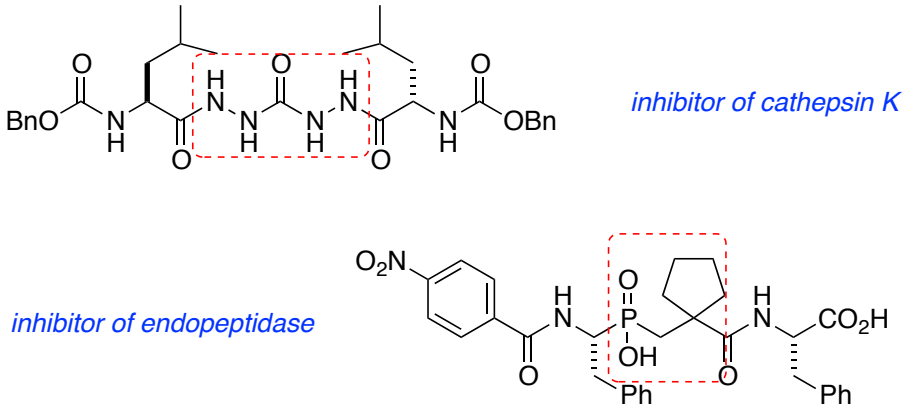
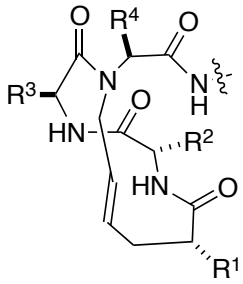
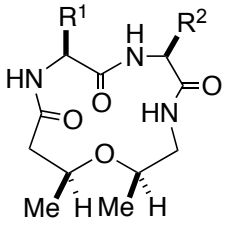


Figure 1.2 (a) Examples of peptide isosteres. (b) Examples of mimics both peptide main-chains and amino acids side-chains.

b



α-helix mimic by Arora



β-turn mimics by Aube

Figure 1.2 continued

Most minimalist peptidomimetics do not have polyamide backbones, but they mimic peptides because of their side-chain compositions and orientations. This approach is legitimate because statistical analyses of structurally characterized protein-protein interfaces show side-chain substituents account for about 80 % of the interactions and main-chain interactions contribute about 11 %.[7] According to this paradigm, many groups have designed analogs of peptide secondary structures that present selected side-chains. In this case, the peptide backbones are replaced with rigid scaffolds.

Figure 1.3 illustrates some examples of minimalist peptidomimetics. Hamilton's group has generated many elegant designs of helical mimicry by using terphenyl scaffolds to present side-chains in appropriate orientations.[8-11] Ahn's group also has designed α-helical mimics based on a bis-benzamide that incorporates four side-chains.[12,13] In β-turn mimicry, Hirschmann and Smith's group reported a series of minimalist β-turn mimics where sugar,[14,15] steroid[16] or catechol[17] backbones were used to hold relevant side-chains. In addition, Arora's group prepared triazolamers as functional β-strand mimetics.[18]

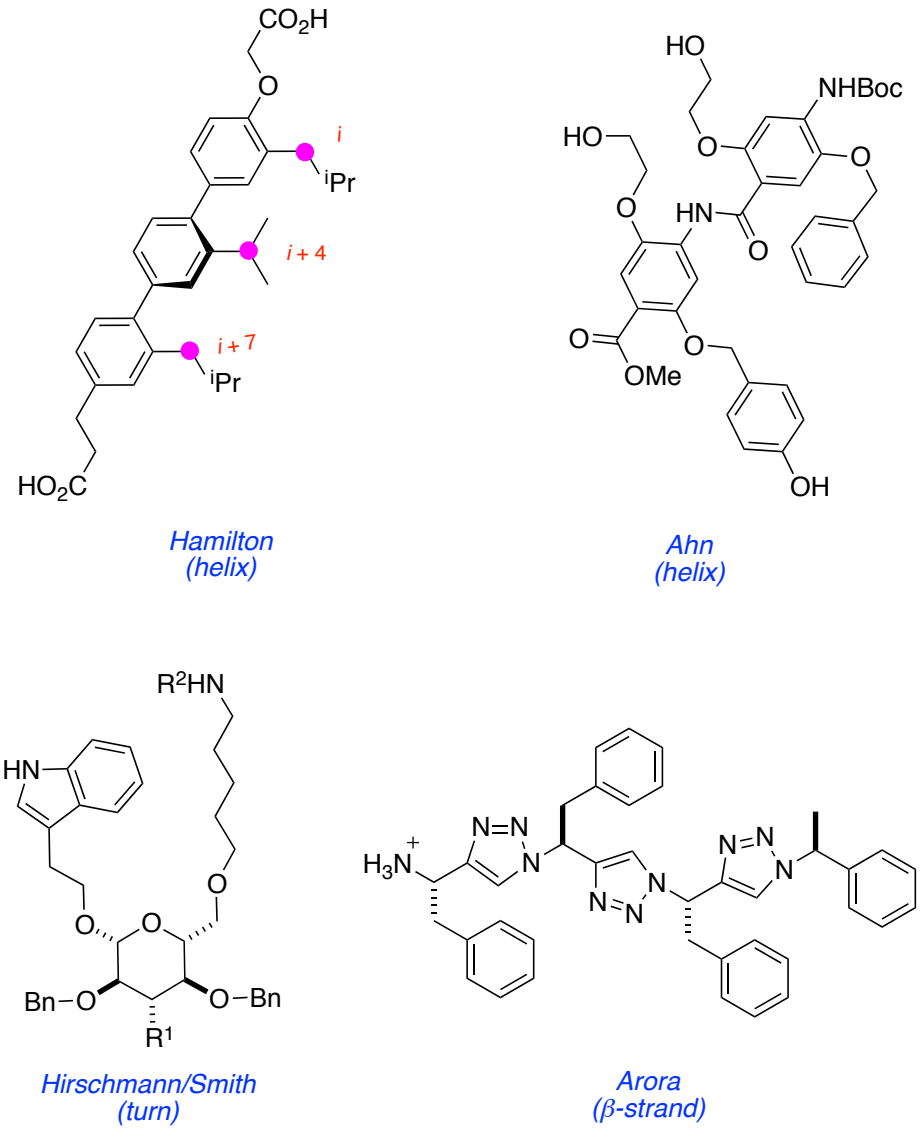


Figure 1.3. Examples of minimalist mimics.

1.2 Methods to Identify Small Molecules that Perturb Predetermined PPIs

It is important to develop small molecules that can disrupt PPIs for the development of pharmacological probes and therapeutics.[7] The most common approaches to do this are high-throughput screening (HTS), fragment-based drug discovery, and computer-aided drug discovery.[19]

1.2.1 High-throughput Screening (HTS)

Since the late 1980s, HTS has been the most widely used approach to find lead molecules for therapeutic applications.[19] An example of drug discovered via HTS is "maraviroc", an antiretroviral that acts as a chemokine receptor antagonist and is used in the treatment of HIV infection (Figure 1.4). UK-107,543 was identified as the most promising hit after HTS of the Pfizer file collection, and then maraviroc was selected from 1,000 analogs via lead optimization processes which included binding potency, antiviral activity, absorption, pharmacokinetics and selectivity against human targets.[20]

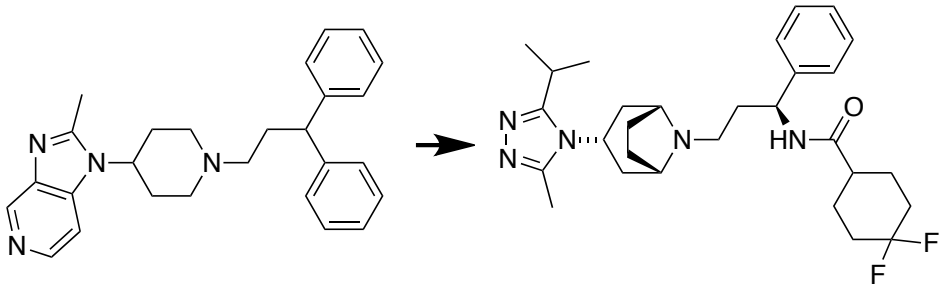


Figure 1.4. Discovery of maraviroc via high-throughput screening.

There are many successful examples of lead compound discovery via HTS.[20] However, HTS has several limitations. First, the success of HTS has not been uniformly distributed among different drug target classes. HTS has been highly successful for traditional targets such as G-protein coupled receptors, ion channels, and enzymes, but much less effective for PPI targets. One of the potential reasons that PPI targets has poor hit rates in HTS is perhaps the compounds were designed for traditional targets. If a library is not chemically suitable for the target, HTS will fail to find a hit. A hypothesis of this thesis is that the wealthy structural information about PPIs could be used to find more developed probes and therapeutics that the ones are currently

emerging from HTS.

1.2.2 Fragment-based Drug Discovery

In fragment-based drug discovery, small molecular fragments (less than MW ~200 – 300 Da) are screened for binding against a target. Fragments that bind to a protein are either elaborated or linked to generate hits with higher affinities (Figure 1.5a).[21] Small fragments tend to be synthetically accessible and more water soluble than larger organic molecules,[22] and less compounds are screened than in classical HTS, so the approach tends to be more facile. An example of fragments-based methods leading to a hit compound is the one from Hannah's group for the immunology target inosine monophosphate dehydrogenase (IMPDH, Figure 1.5b).[23-25]

A problem with HTS that is accentuated in fragment-based approaches based on binding is that it tends to be necessary to screen high concentrations of compounds to find weak binders. High concentrations are impractical for many compounds because of solubility and quantity considerations. Moreover, high concentrations often result in false positives.[22]

a

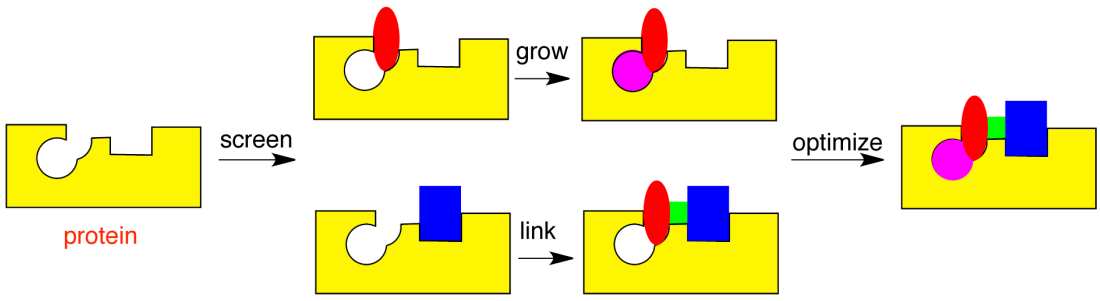


Figure 1.5. Fragment-based lead discovery: (a) basic approach[21]; and, (b) an example applied to an immunology target, inosine monophosphate dehydrogenase.

b

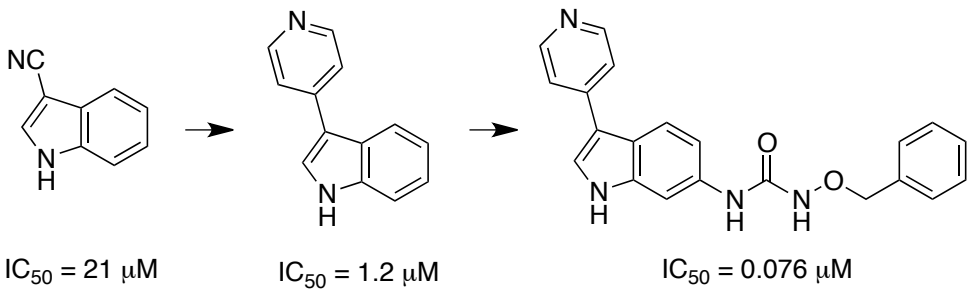


Figure 1.5. continued.

A subset of fragment-based methods is Wells' "tethering" approach that largely overcomes the high concentration issue. This strategy uses the reversible covalent bond formed between a disulfide-linked fragment and a cysteine residue near the binding site in a target protein as illustrated in Figure 1.6.[22] Thiol-disulfide exchange amplifies the affinity of the fragment for the target protein, and it enables use of low concentrations. Mutagenesis allows cysteines to be incorporated at specific locations for added control over regions where fragments bind the protein.

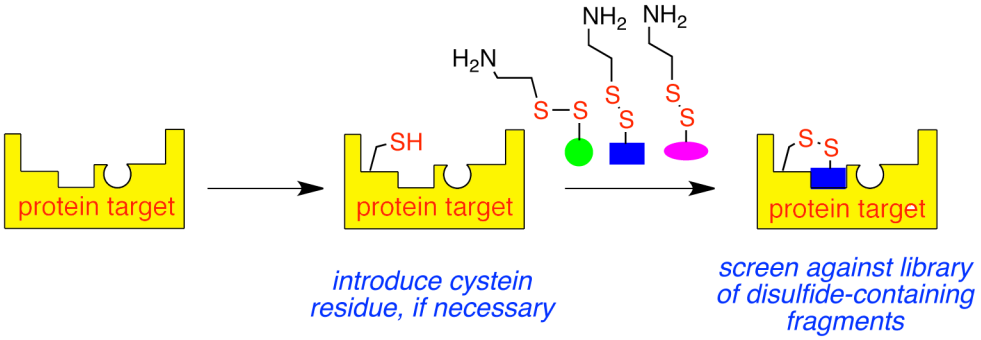


Figure 1.6. Tethering for fragment assembly.[22]

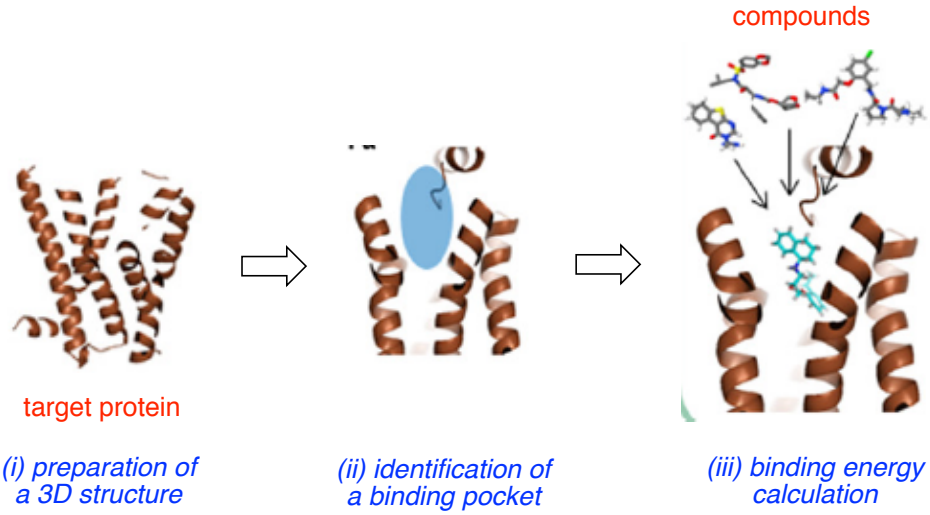
However, it is impossible exactly to predict the binding affinities of fragments for targets. Other fragment-based approaches to overcome this limitation are to use structural methods, NMR spectroscopy[26] or X-ray crystallography[27]. NMR scpectroscopy uses NMR chemical shifts after a ligand binds to a protein, and X-ray crystallographic method monitors changes of the electron density map of a complex relative to the unbound form.

1.2.3 Computer-aided Drug Discovery

Computer-aided drug discovery involves simulation of drug-target interactions using tools for bioinformatics and databases to filter undesired molecules from libraries, even before synthesis,[28] reducing expenditure on resources and time. There are two main paradigms in computer-aided methods: structure- and ligand-based approaches.

Structure-based approaches use target protein structures and docking algorithms to place computer-generated representations of a small molecule;[29] this method has been particularly successful for enzyme active sites (Figure 1.7a).[30] The word "pose" describes each docking mode. Poses are evaluated ("scored") based on the ability of the small molecule to complement the target physicochemical properties (e.g. shape and electrostatics). This process is repeated for all molecules in the library, and the results are rank-ordered by scores to select compounds for further investigation (Figure 1.7b).[29]

a



b

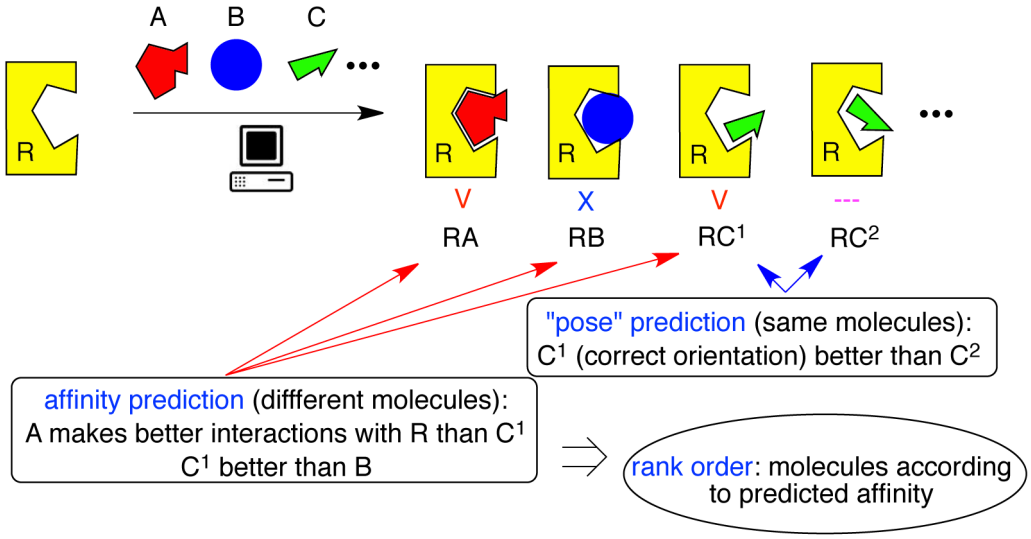


Figure 1.7. (a) Identification of binding sites on a target protein[30]; and, (b) docking and scoring. R = a receptor protein, A, B, and C = small molecules to be docked in the receptor.[29]

Computer-aided drug design relies on high-resolution protein structure and *good justification for focusing on certain narrowly defined regions where the small molecules can bind*.[19] PPI interfaces between proteins are relatively extensive and there can be many thermodynamically preferred sites for binding a small molecule, including ones that are not at the interface. Consequently, even if the thermodynamically significant interface regions (hot-spots) for binding the other protein are known (and in most cases they are not), then it is perfectly possible that a small molecule would bind at another site on the proteins, if it binds at all. For these reasons computer-aided drug design is not ideal for finding small molecules that perturb protein-protein interactions.

Ligand-based methods use known ligands as templates to discover new binding molecules.[19,31] Figure 1.8[30] indicates how known active ligands are classified according to their physicochemical properties (molecular weight, logP, polar surface area, *H*-bond donor and acceptor counts); parameters such as 2D (scaffold, substructure, fingerprints) or 3D (pharmacophore) descriptors are stored in a database, and then screening is performed via quantitative structure-activity relationship (QSAR) analyses, machine learning (Bayesian classification, support vector machines) or pharmacophore searches. New potential ligands are then conceived and screen *in silico* to identify ones that match the physicochemical properties of the reference compounds.[32] Obviously, this approach requires good known ligands;[33] it cannot be applied to *de novo* discovery of unexplored PPIs.

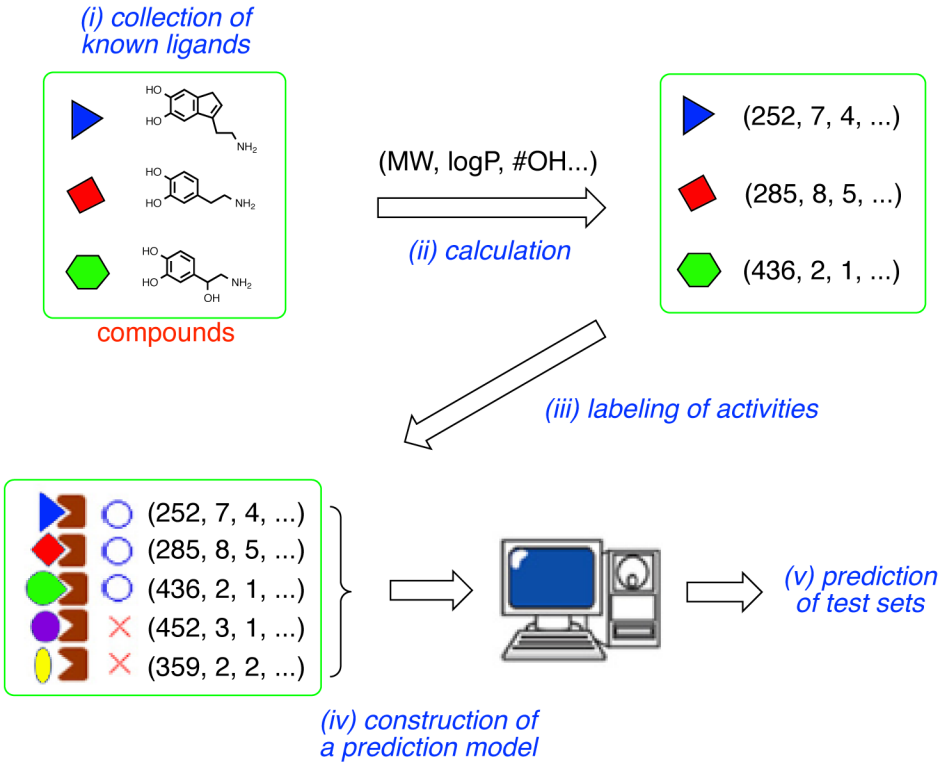


Figure 1.8. Ligand-based computational methods for drug discovery.[30]

1.3 Conclusion

PPIs are important targets because of their roles in cell signaling and their therapeutic potential. There is abundant, detailed structural information about interfaces between proteins, but state-of-the-art methods for finding small molecules that impact PPIs *cannot* use this very effectively. Many of those methods consider only general trends in the structural information, e.g. prevalence and "druggability" of helical regions.[5] Concepts like "minimalist peptidomimetics" are interesting, but at the onset of our studies there was no way to relate them accurately to specific structural data; crude overlay procedures onto known ligands marks an intellectual boundary in this work. Moreover, both types of computer-aided drug discovery methods, structure- and ligand-based, are *not* well suited to PPIs. Experimental fragment-based procedures that feature NMR or X-ray use structural information about pre-selected targets, but fragment-based methods based on binding assays do not. Consequently, there is much structural data available that cannot be used in fragment-based methods. For example, there are severe restrictions on NMR methods to be used with huge proteins, and X-ray approaches cannot be used if crystals of a protein are not amenable to, or available for, soaking experiments. Conventional HTS is probably the crudest approach; it does not make use of structural information at all.

My Ph D research began by focusing on very specific protein-protein interactions (neurotrophin-Trk), then gradually turned towards closer consideration the salient features of minimalist mimics. Finally, this led us to focus on the question with the broadest impact in this field, specifically, *are there better ways to use protein-protein interaction structural data to identify small molecule pharmacological probes and pharmaceutical leads*.

CHAPTER II

DESIGN OF TRIAZOLE-BASED MIMICS TO TARGET Trk RECEPTORS

2.1 Introduction

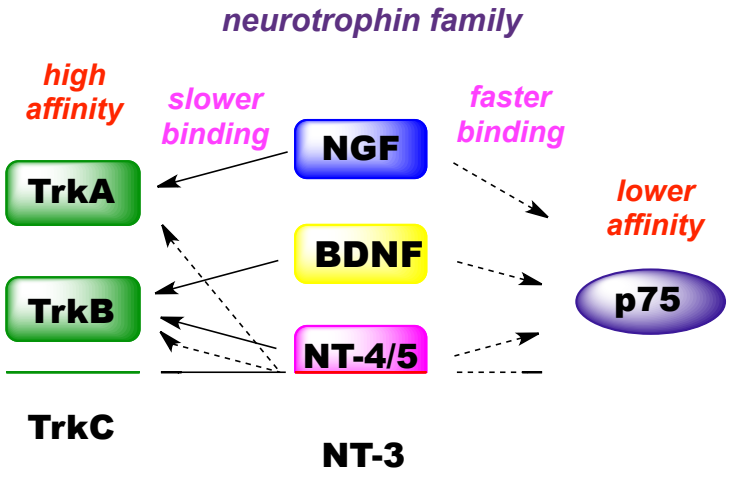
Neurotrophins are a family of growth factors that play key roles in the development and maintenance of the peripheral and central nervous systems (PNS and CNS).[34,35] The family includes nerve growth factor (NGF), brain-derived neurotrophinc factor (BDNF), neurotrophin-3 (NT-3), and neurotrophin-4 (NT-4). These proteins bind to the ***t***ropomyosin-***r***eceptor-***k***inase (Trk) receptors, (TrkA, TrkB, and TrkC) and the pan-neurotrophin receptor p75 with different binding affinities. The high affinity binding interactions are for NGF with TrkA, BDNF and NT-4 with TrkB, and NT-3 with TrkC; somewhat lower affinity, non-selective, binding is observed for all ligands with p75 (Figure 1a).[36,37] Binding of neurotrophins to Trk receptors leads to activation of tyrosine kinases, receptor dimerization, phosphorylation, and downstream signaling {phosphatidylinositol 3-kinase (PI3K)/Akt, mitogen activated protein kinase (MAPK), and phospholipase C-γ1 (PLCγ) pathways}. Neurotrophins signaling through Trk receptors regulates cell survival, proliferation, growth, regulation of channel functions, and assembly of cytoskeleta (Figure 1b).[34,37]

The p75 receptor is a member of the tumor necrosis factor (TNF) receptor superfamily. Its functions are less clearly understood than they are for Trk receptors. Binding of neurotrophins to p75 receptor can lead to apoptosis or survival depending on the cells used and the stage that they are at their life cycles (Figure 1b).[34]

Activation of Trk and p75 receptors can lead to *opposite* effects on cells, e.g. growth/differentiation and apoptosis,[35] thus a major barrier to using neurotrophins as drugs is the fact that they can activate both receptors. On the other hand, small

molecules that *only* bind and activate Trk receptors could be extremely useful as drugs and pharmacological probes.

a



b

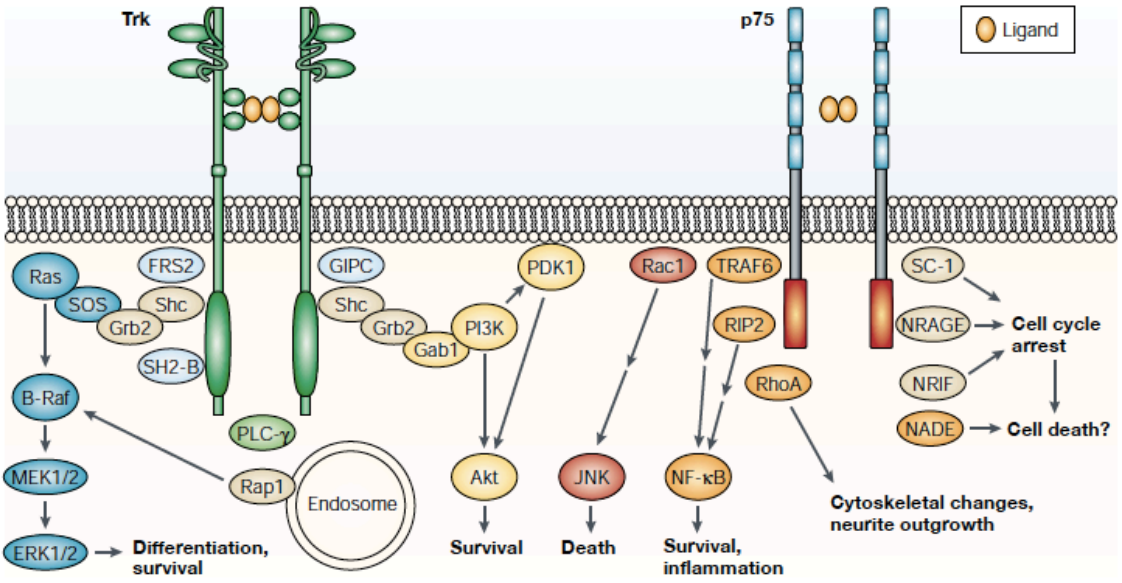
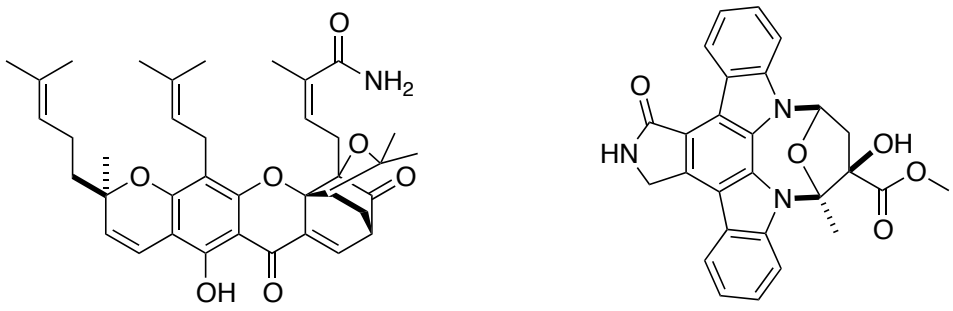


Figure 2.1. (a) Affinities of neurotrophins for Trk and p75 receptors. (b) Neurotrophin receptors signaling.

Neurotrophins and Trk receptors are validated targets to treat a variety of pathologies ranging from neurodegenerative diseases to cancers.[35,38-40] In most cells, activation of Trk receptors by neurotrophins results in survival and neurite outgrowth responses.[37] NGF supports dorsal root ganglion (DRG) neurons during development and small fiber nocieptors in the adult. NT-3 promotes survival of adult proprioceptive DRG neurons and cholinergic neurons, and BDNF supports large fiber DRG neurons and cholinergic neurons.[41] These critical roles of neurotrophins during development and in the adult nervous system mean that agonists for Trk receptors might be useful for the treatment for neurodegenerative diseases and stroke. For example, gambogic amide (Figure 2.2a) selectively binds to TrkA, induces receptor dimerization, tyrosine phosphorylation, and Akt and MAPK signaling and it is therefore a potentially useful probe.[42] This compound acts as a TrkA agonist and prevents neuronal apoptosis initiated by kainic acid (KA; a compound that induces neuronal cell death in caspase-dependent and caspase-independent pathways).

On the other hand, Trk receptors are activated in a broad range of cancers, where they modulate tumor growth and motility.[39] Antagonists for Trk receptors might provide effective treatments for cancers. For instance, K252a in Figure 2.2b is an alkaloid isolated from *Nacardiopisis* sp. Siol fungi. It blocks neuronal differentiation of rat pheochromocytoma PC12 cells induced by NGF. Moreover, it is a potent inhibitor of the tyrosine protein kinase activity of the NGF receptor, gp 140trk that is the product of the trk protooncogene.[43]

a **b**



gambogic amide

agonist for TrkA

K252a

antagonist for Trks

Figure 2.2. (a) Gambogic amide, a selective agonist for TrkA. (b) K252a, a selective inhibitor for Trk receptors.

Neurotrophins are noncovalent homodimers of *ca* 120 amino acid protein units. Each monomer has three solvent-exposed β-turn regions[35] that seem to be "hot-spots" for interaction between neurotrophins and Trk receptors.[36] Table 2.1 is a summary of the hot-spot residues in β-turn regions of neurotrophins. In small molecules that mimic neurotrophins, selectivities for TrkA, B and C might be achieved by amino acid side-chains in β-turn regions of each neurotrophins.

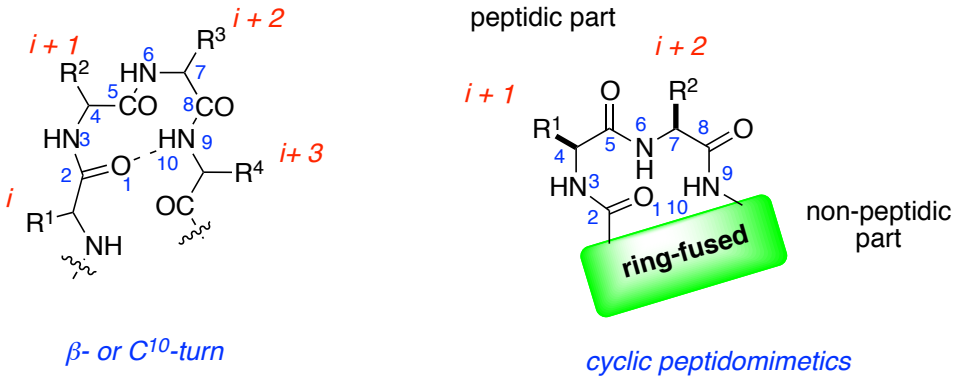
Table 2.1. Summary of "Hot-spot" Residues in Neurotrophins.

proteins	source	β-turn sequence		
		30-33	**44-47**	**93-96**
NGF	murine	DIKG	INNS	DEKQ
	human	DIKG	INNS	DGKQ
	bovine	DIKG	INNS	DNKQ
	guinea pig	DIKG	VNNN	DGKQ
		30-33	**44-47**	**93-96**
BDNF	pig	DMSG	VSKG	DSKK
	human	DMSG	VSKG	DSKK
		29-32	**42-45**	**92-95**
NT-3	mouse	DIRG	KTGN	ENNK
	human	DIRG	KTGN	ENNK

[a] *i*+1-*i*+2 sequences that are highlited with red were chosen for syntheses in this study.

Our group has designed and synthesized small molecules that can mimic the β-turn hot-spots.[44-49] The first generation structures are cyclic peptidomimetics, such as D3.[44-46] This compound was designed to mimic *i*+1 and *i*+2 of the β-turn in NGF through "ring fused C^{10}-motif" as Figure 2.3a and b.[45] D3 is a partial agonist for TrkA; the compound induced cell survival, neurite outgrowth, and tyrosine phosphorylation *in vitro* assays.[50] In addition, the small molecule caused recovery of loss of memory function in an *in vivo* an "aged rat model".[51]

a



b

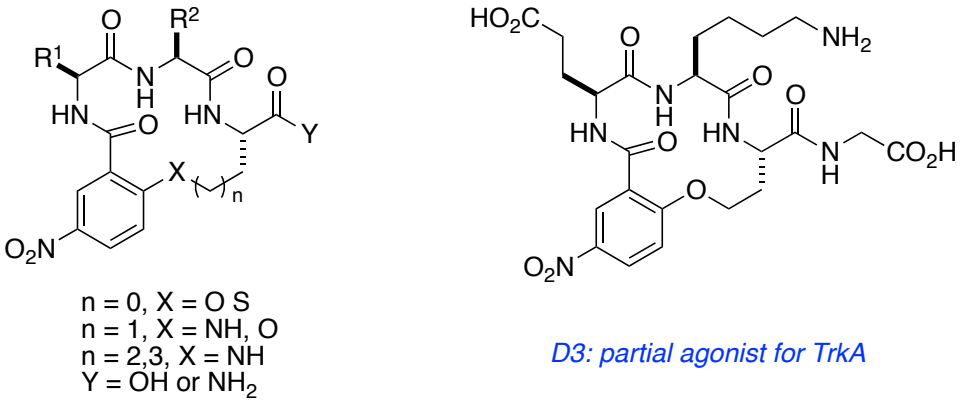


Figure 2.3. β-Turn cyclic peptidomimetics with ring-fused C^{10} motif.

Second generation structures developed by our group are small molecules, "minimalist mimics", that express only side-chains of the β-turn part and have no molecular fragments to resemble the main-chain.[52] For the structure in Figure 2.4a, a heterocyclic scaffold is used instead of peptide backbone. The scaffold gives rigidity to limit the degrees of freedom for bond rotations but, ideally, without excluding conformations that correspond the targeted secondary structures. The first minimalist mimics in our group were built on triazole-based scaffolds (Figure 2.4b).[47] The triazole core was obtained by click chemistry between azido amino acids and alkyes having side-

chains of amino acids (Figure 2.4c). The triazole-based β-turn mimics have piperazine linkers to build bivalent peptidomimetics that are obtained by S_NAr reactions on triazines. Combinatorial methods shown in Figure 2.4d were used to combine two different monovalent units to give bivalent mimics (hetero- or homo-). This combinatorial method has several advantages. First, combination of *n* monovalent compounds gives $n(n+1)/2$ bivalent compounds; a *lot* of bivalent mimcs can be made from relatively few monovalent ones. Second, the products can be formed without using any protecting group on monovalent components. Third, the third site in triazine core can be used to support tags for biological assays, e.g. dyes, biotin, and, as described in the next chapter, cytotoxic compounds. Before I began this effort, Dr Yu Angell, a preceeding student in our group, had prepared 78 bivalent peptidomimetics with fluorescein tags and had them tested for selective binding to Trk receptors (in fluorescence-activated cell sourting, FACS, assays). Four of the peptidomimetics she made (Figure 2.4e) showed preferential binding to TrkA-expressing cells.

a

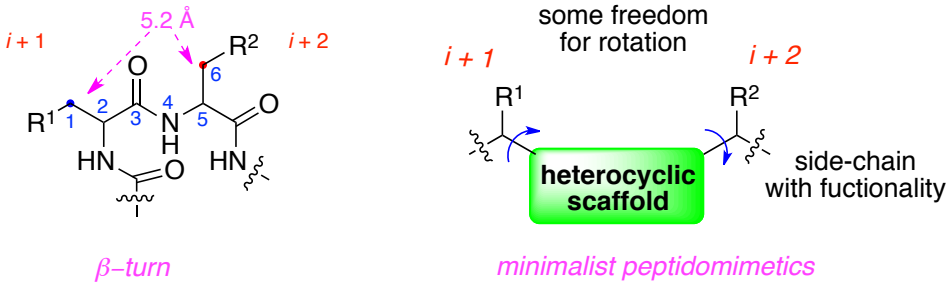
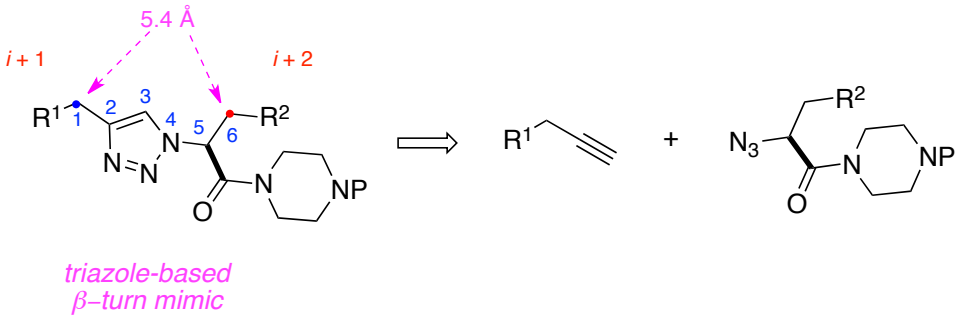


Figure 2.4. (a) Design of minimalist peptidomimetics with triazole scaffolds. (b) Overlay of triazole-based mimic with type I β-turn. (c) Strategy to build triazole-based β-turn mimics. (d) A general scheme to prepare bivalent mimics. (e) Structures of compounds that bind to TrkA.

b

c



d

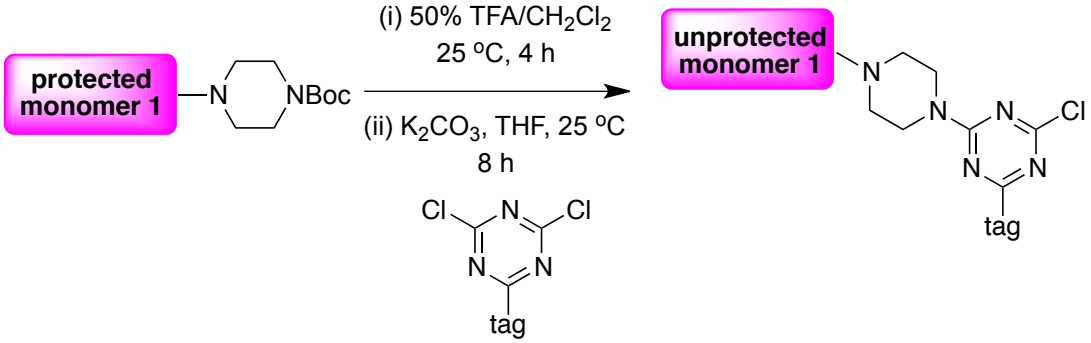


Figure 2.4. continued.

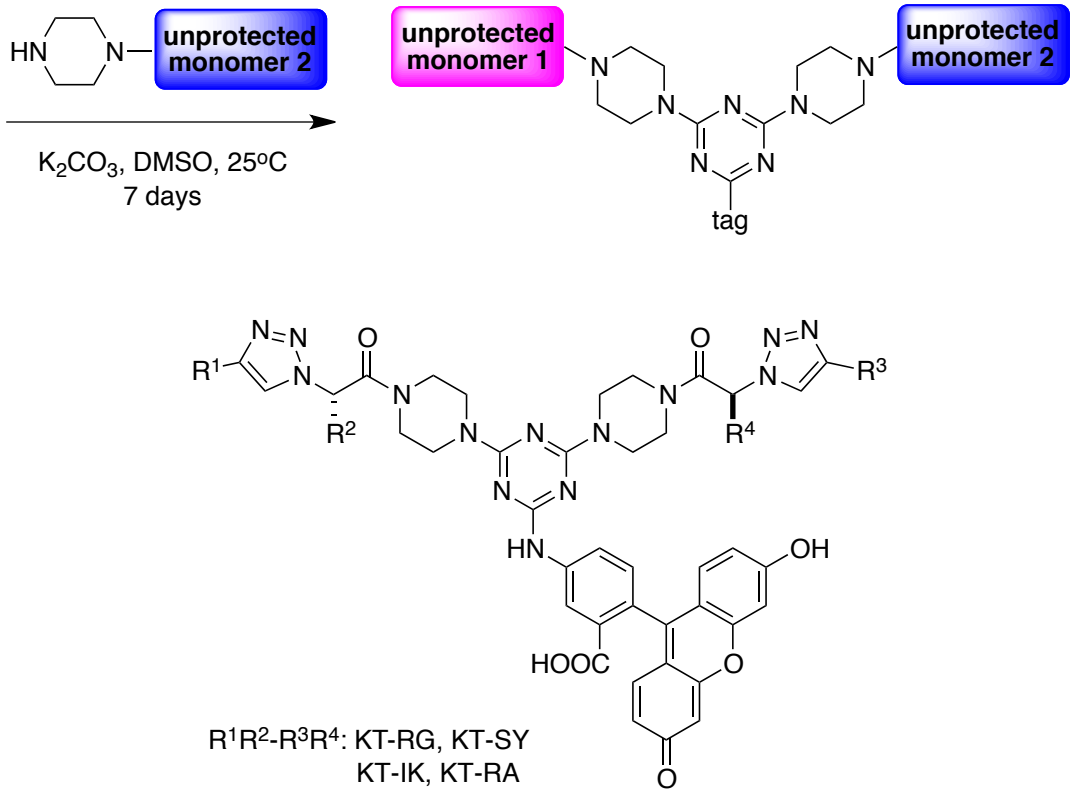
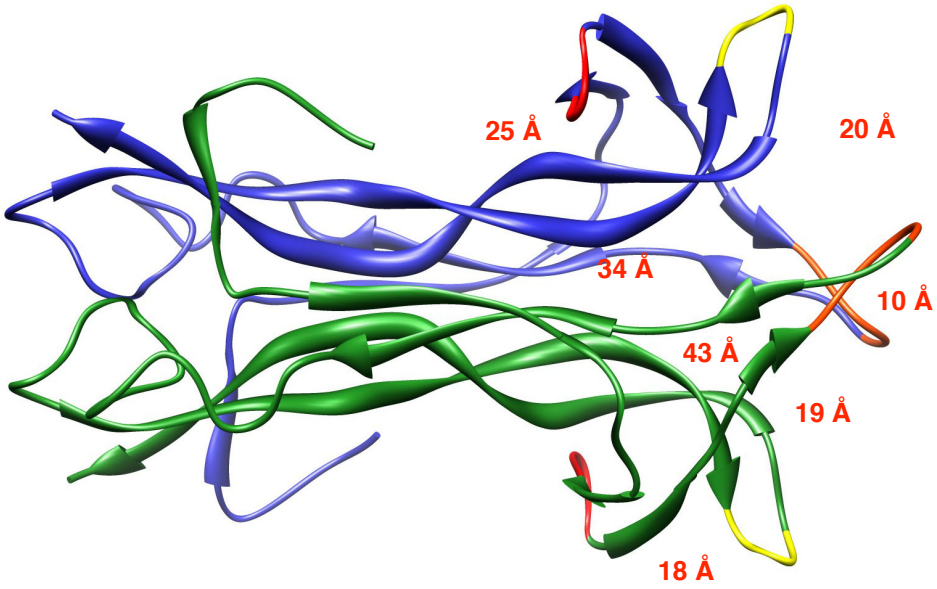


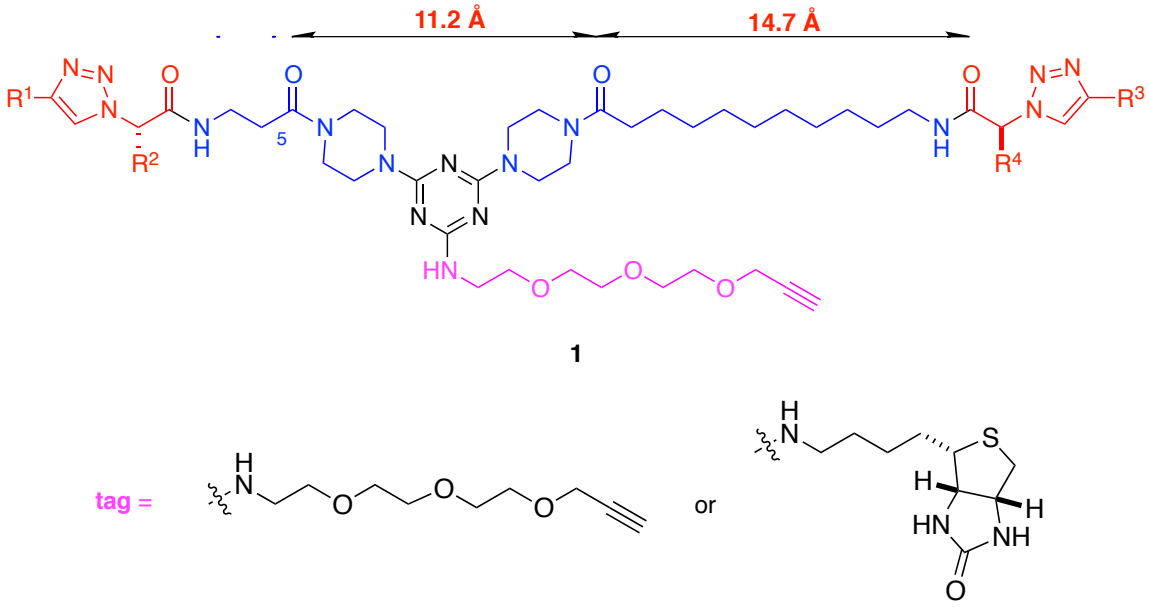
Figure 2.4. continued.

The results described above proved that the triazole-based compounds can bind TrkA expressing cells just as NFG does, but their activities in cellular assays was weak. My role was to discover related compounds to improve binding to Trk receptors and the activities of these compounds in cellular assays. In particular, we were interested to use long-linkers between triazole core and triazine core in bivalent mimics to provide more distance variation for separation of the monovalent units on binding. Distances between hot-spots in neurotrophins are from 10 Å to 43 Å (Figure 2.5a). Consequently, we reasoned that if each monovalent mimic were connected to flexible chains of about 15 Å length, the total separation would be about 41 Å maximum, and the warhead regions could orient like *any* pair of turns in NGF (Figure 2.5b). This strategy might especially increase the chances for mimicry of β-turns that are far from each other.

a



b



1

Figure 2.5. (a) Distances between hot-spots in NT-3. (b) Structures of triazole-based β-turn mimics with long linkers, and lengths of triazine core and linker. (c) Strategy to build bivalent mimics with different orientation sequences in triazole core.

c

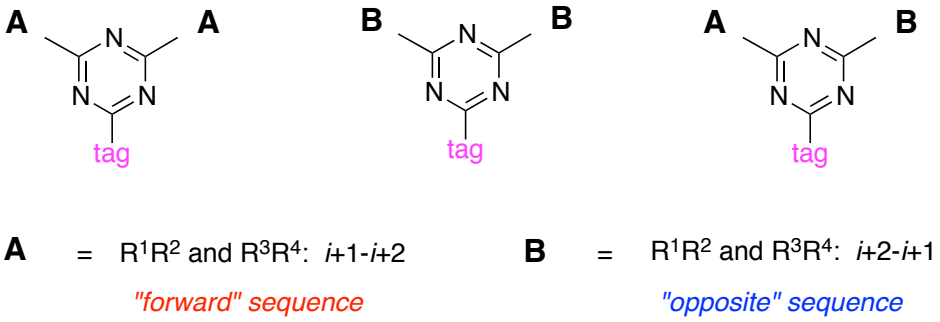


Figure 2.5. continued.

Another new line of investigation for my studies was to use side-chains on the peptidomimetics that correspond exactly to the *i*+1 and *i*+2 residues in the neurotrophins turn regions (Table 2.1, red highlight). Dr Yu Angell had not done this thoroughly; the residues that she used tended to be selected because they were commonly found at hot-spot interactions in general. Thus, in my work, seven dipeptide sequences were targeted; IK, EK, and GK from NGF, MS and SK from BDNF, and IR and TG from NT-3.

A third aspect that our group had not previously considered carefully the *orientation* of the side-chains in the mimics as illustrated in Figure 2.5c. Monovalent mimics with the long linkers, **1** were prepared in two different orientations ("forward sequence"; R^1R^2 and R^3R^4 = *i*+1-*i*+2, e.g. TG, "opposite" sequence; R^1R^2 and R^3R^4 = *i*+2-*i*+1, e.g. GT), and bivalent mimics by combination of the monovalent mimics were prepared in three different sets (forward only; R^1R^2 and R^3R^4 = *i*+1-*i*+2 and *i*+1-*i*+2, e.g. TG-TG, opposite only; R^1R^2 and R^3R^4 = *i*+2-*i*+1 and *i*+2-*i*+1, e.g. GT-GT, and forward-opposite; R^1R^2 and R^3R^4 = *i*+1-*i*+2 and *i*+2-*i*+1, e.g. TG-GT) as Figure 2.5c.

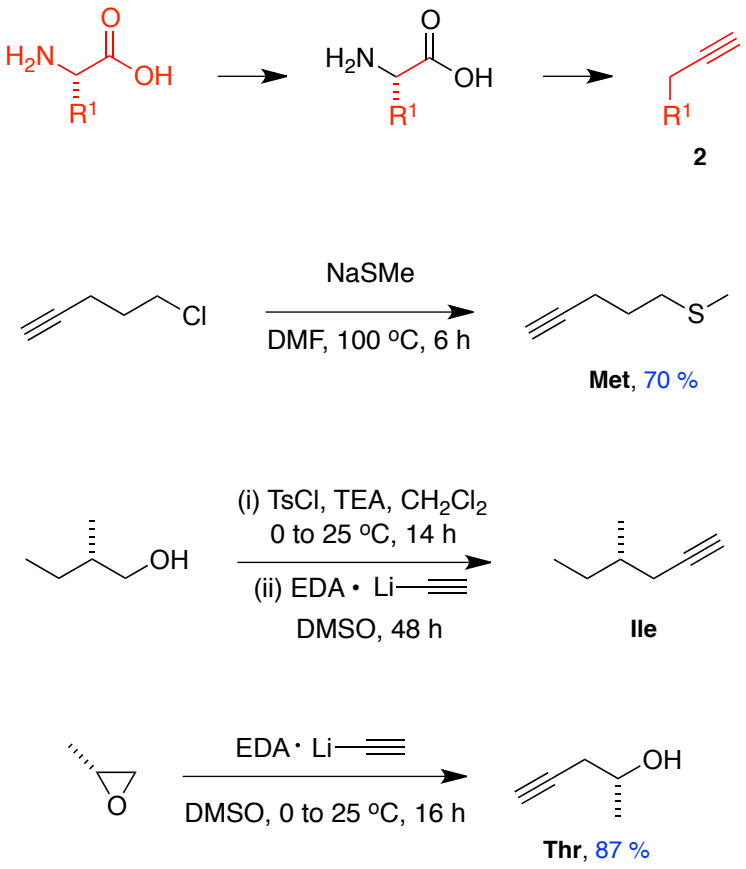
2.2 Syntheses of Triazole-based Monovalent Mimics with Long Linker

The triazole-based peptidomimetics have amino acid side-chains corresponding to *i*+1-*i*+2 of β-turn regions in neurotrophins. Alkynes with the amino acid side-chains

were prepared as Scheme 2.1a. Alkynes **2** corresponding to Glu, Gly, and Ser are commercial available, and those corresponding to Met, Ile, and Thr were obtained by S_N2 reaction (Scheme 2.1.a).[53-55] Azido amino acids **3** for Arg, Gly, Lys, and Ser were prepared via azo-transfer reactions (Scheme 2.1.b).[56] To prepare a long alkyl chain linker, Boc-piperazine was connected to bromoundeconoic acid through amide coupling, and then the primary amine **6** was prepared via Gabriel synthesis.[57] Azido amino acid **3** was connected to the linker through amide coupling, and then triazole core **8** was obtained with the azide **7** and alkyne **2** via Cu(I)-mediated Click reaction.[58]

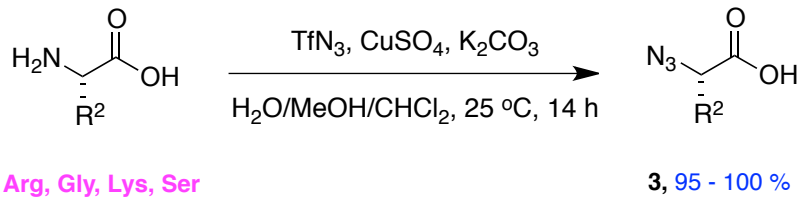
Scheme 2.1. Preparation of Monovalent Mimics with Long Linkers.

a

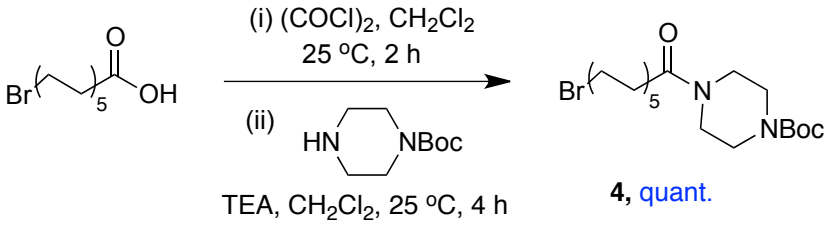


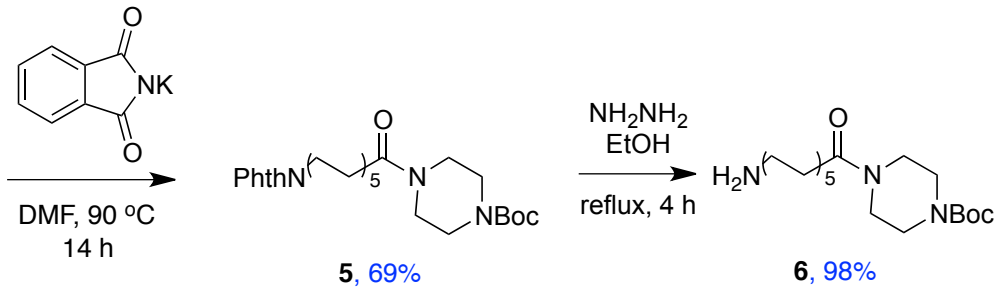
Scheme 2.1. continued.

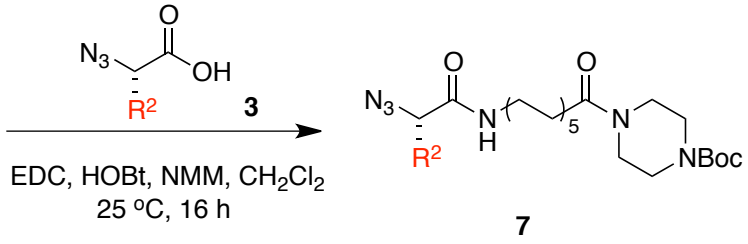
b



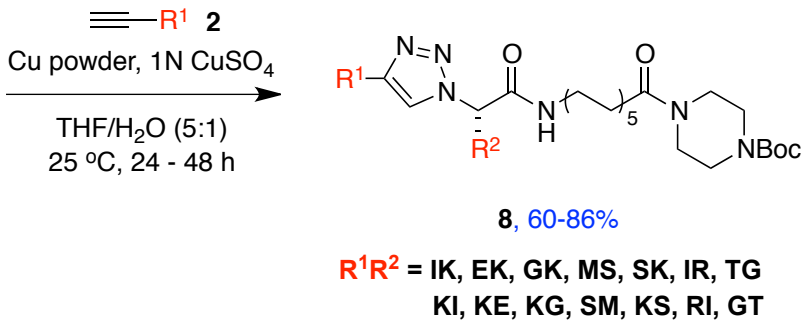
c







Scheme 2.1. continued.

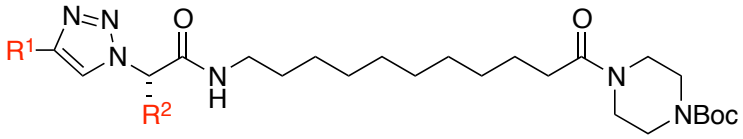


8, 60-86%

R^1R^2 = IK, EK, GK, MS, SK, IR, TG
KI, KE, KG, SM, KS, RI, GT

Characteristics of the seven triazole-base monovalent mimics **8** with orientation of R^1R^2 = *i*+1-*i*+2 were summarized in Table 2.2. Seven monovalent mimics with corresponding but opposite orientations (R^1R^2 = *i*+2-*i*+1) were prepared by Dr Andrey Malakhov.

Table 2.2. Monovalent Peptidomimetics **8a - f**.



8

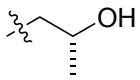
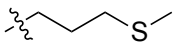
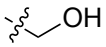
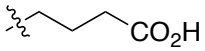
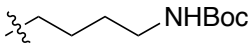
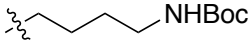
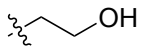
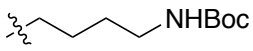
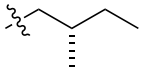
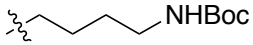
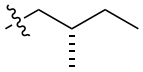
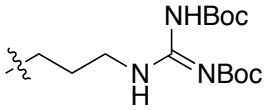
compounds	$R^{1'}$	$R^{2'}$	yield (%)
8a		H	61
8b			60
8c			83
8d			86

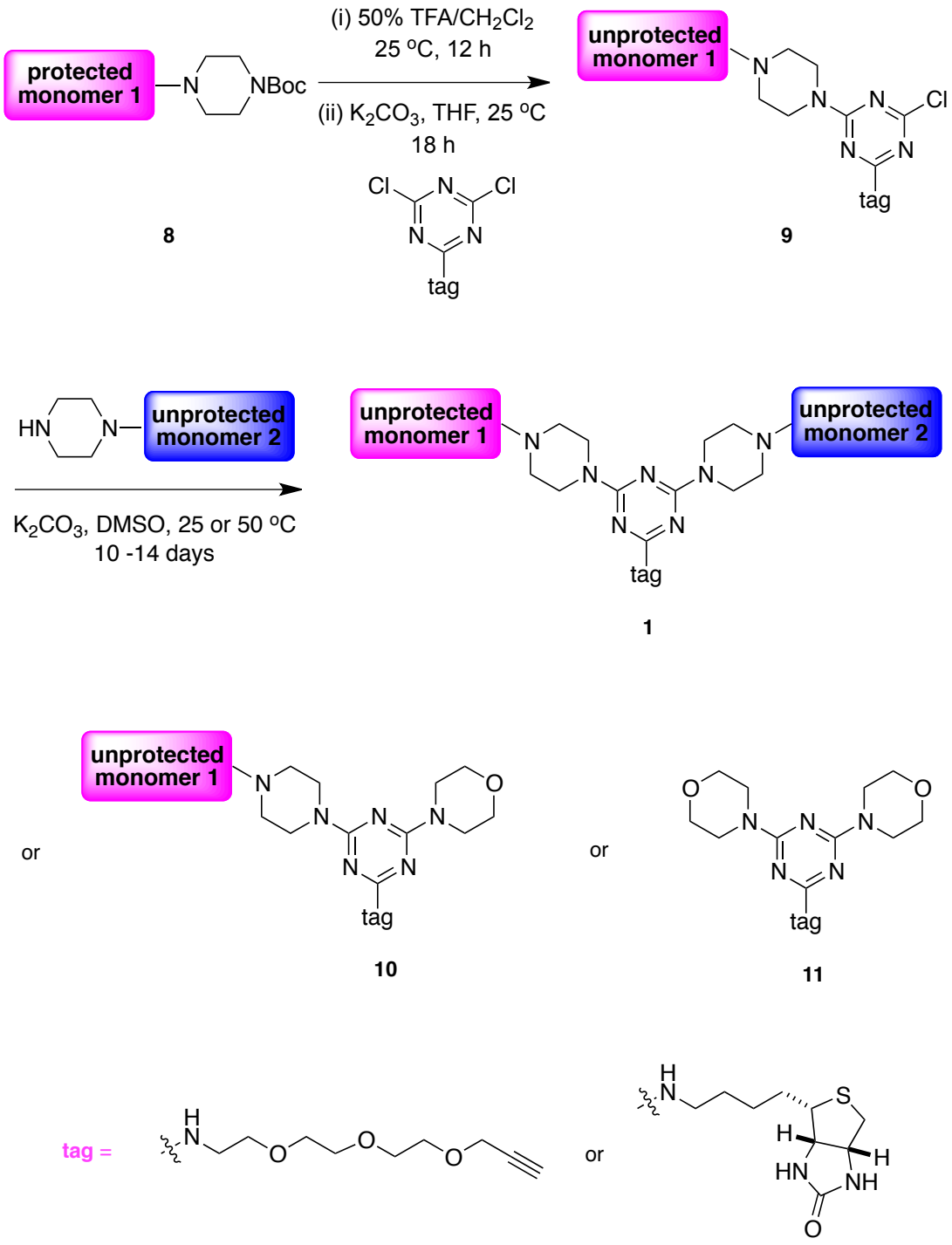
Table 2.2. continued.

compounds	$R^{1'}$	$R^{2'}$	yield (%)
8e			65
8f			81
8g			60

2.3 Syntheses of Bivalent Mimics

A library of 120 bivalent compounds **1** with TEG and six bivalent compounds with biotin were prepared from 14 monovalent compounds and morpholine using a similar procedure to the triazole-based bivalent mimics prepared by Dr Yu Angell (Scheme 2.2).[47] The monovalent mimics **8** were deprotected with TFA, and then the deprotected compounds were added to solution of equimolar amounts of cyanuric chloride with TEG or biotin tag and excess K_2CO_3 in THF. The reactions were monitored by analytical HPLC. After the reactions were done, THF was removed, and then the second deprotected monovalent mimic was added to the intermediates **9** in DMSO, and excess K_2CO_3. The solution was stirred at 25 or 50 °C for more than 10 d to get the bivalent mimics **1**. After monitoring the reaction in HPLC, the crude compounds were lyophilized to remove DMSO, and then purified with RP-preparative HPLC to obtain the final bivalent mimics. Purities of 77 bivalent mimics **1**, seven bivalent mimics with morpholine cap like **10** and **11**, and seven deprotected monovalent mimics (NH_2) were summarized in Figure 2.6. Full detail of analyses materials is given in Appendix B.

Scheme 2.2. Preparation of The Tagged Bivalent Peptidomimetics **1**.



a.

	8a	8b	8c	8d	8e	8f	8g
8a[a]	88						
8b	97	100					
8c	98	91	90				
8d	93	100	89	100			
8e	98	95	99	98	98		
8f	100	93	94	98	94	98	
8g	100	99	99	87	98	100	100
cap[b]	91	100	98	95	91	100	95
NH_2[c]	94	100	99	99	97	90	96

UV purities (%)
90 and above
Below 90

b.

	8a	8b	8c	8d	8e	8f	8g
8a[a]	100						
8b	100	100					
8c	100	95	99				
8d	100	100	90	100			
8e	100	100	100	95	99		
8f	100	100	100	99	99	100	
8g	100	100	98	96	100	100	100
cap[b]	96	100	100	100	100	97	100
NH_2[c]	100	100	100	100	99	100	100

SEDEX purities (%)
90 and above

Figure 2.6. Purities of the library of compounds **1**. (a) UV detection; and (b) SEDEX detection after purification for a natural orientation library. (c) SEDEX detection after purification for a mix orientation library. [a] Deprotected forms of **8**, [b] Cap=morpholine, [c] Deprotected monovalent mimics' purities.

c.

	8a	8b	8c	8d	8e	8f	8g
8a	100	100	100	100	100	100	100
8b	100	96	100	100	100	100	100
8c	87	100	100	99	100	100	99
8d	100	100	100	100	100	100	100
8e	100	99	100	97	100	100	100
8f	100	100	100	100	100	100	100
8g	100	100	100	100	100	100	100

SEDEX purities (%)
90 and above
Below 90

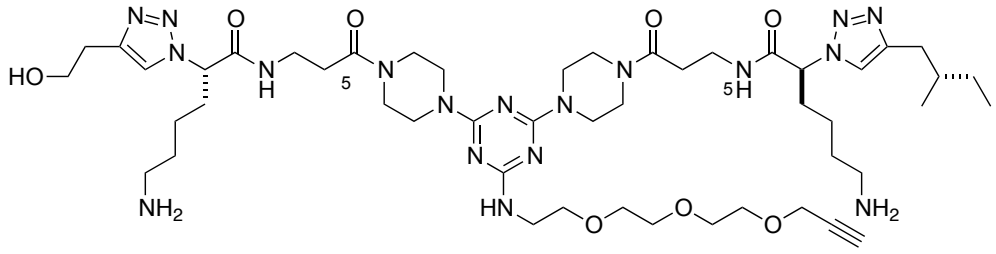
Figure 2.6. continued.

2.4 Biological Assays

Biological assays were conducted by Dr Uri Saragovi and his coworkers at McGill University in Canada. Cell survival assays,[59] and signal transduction assays[60] were carried out for the TEG-labeled bivalent mimics and deprotected monovalent mimics. A direct binding assay[61] was carried out for biotin-labeled bivalent mimics. For the assays, NIH-3T3 cells were transfected with each Trk receptor; NIH-TrkC cells express ~100,000 TkrC receptors/cell, NIH-TrkA cells express ~200,000 TrkA receptors/cell, and NIH-IGF-1R cells express ~100,000 IGF-1R receptors/cell. Parental NIH-3T3 cells are mouse fibroblasts that do not express any neurotrophin receptors. Neuronal PC12 cells express TrkA and p75 neurotrophin receptors and respond to NGF, and nnr5-TrkC cells are variant of PC12 that lack TrkA and into which human TrkC cDNA was stably transfected and respond to NT-3. The 4-3.6 cells are B104 rat neuroblastoma stably transfected with human TrkA cDNA and express equal levels of p75 and TrkA.[62]

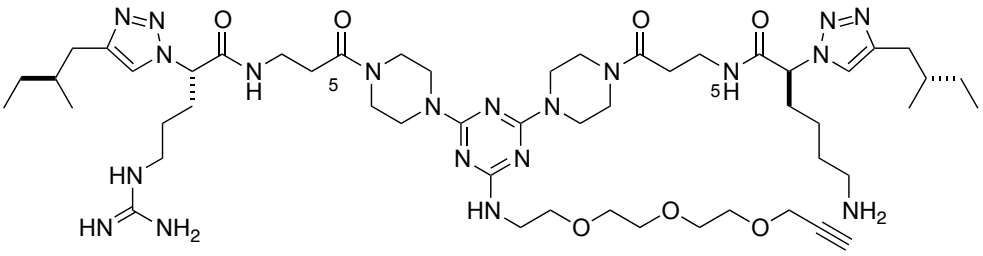
2.4.1 Cell Survival Assays

A library of 120 bivalent compounds including morpholine cap control compounds and 14 deprotected monovalent compounds were tested for cell survival to transfected Trk receptor cells. The cell survival was measured quantitatively by the MTT assay in 96-well plates. The peptidomimetics were tested at non-cytotoxic doses for their effect on NGF or NT-3 mediated survival. From the screening, six compounds **1ef**, **1gf**, **1f'e'**, **1a'c**, **1g'a**, and **1b'd** in Figure 2.7 reduce the survival of NIH-TrkC cells responding to NT-3 (Figure 2.8a). Two of them **1ef**, and **1gf** block the survival of TrkA induced by NGF, as well (Figure 2.8b). In control tests, none of the compounds showed any significant survival effect for NIH-IGF-1R cells. Similar results were observed in p75 co-expressing cells, nnr5TrkC cells, and PC 12 cells (Figure 2.8c). p75 co-expression does not interface in the antagonism of survival by the compounds.

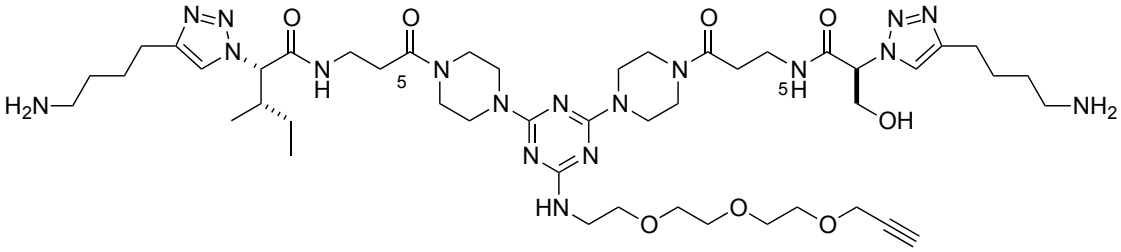


1ef
TrkC and TrkA antagonist

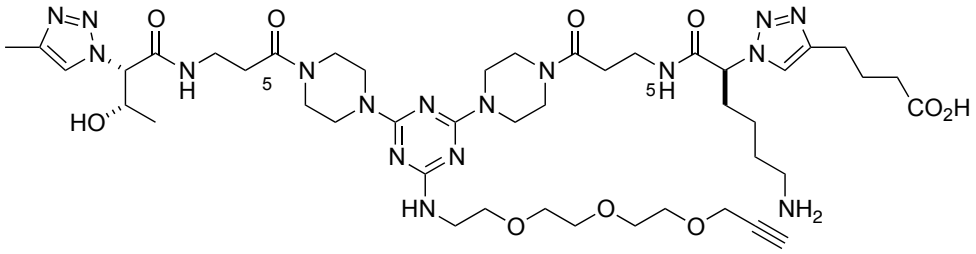
Figure 2.7. Structures of antagonists for TrkC or TrkA. Prime on labels indicated the opposite sequences (*i*+2-*i*+1) of sequences in Table 2.2.



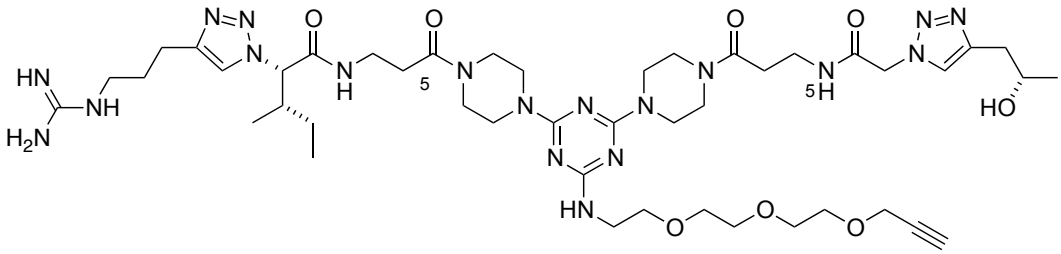
1gf
TrkC and TrkA antagonist



1f'e'
TrkC antagonist

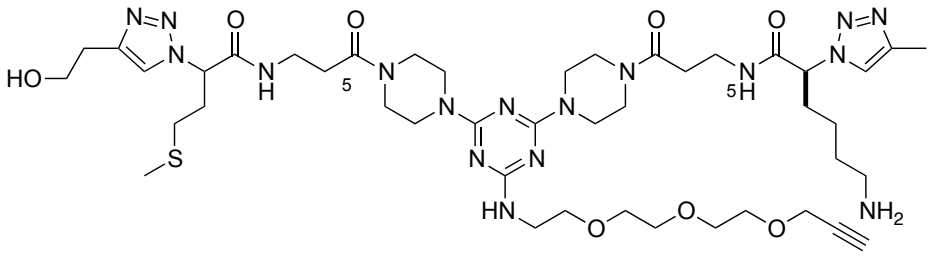


1a'c
TrkC antagonist



1g'a
TrkC antagonist

Figure 2.7. continued.



1b'd
TrkC antagonist

Figure 2.7. continued.

a

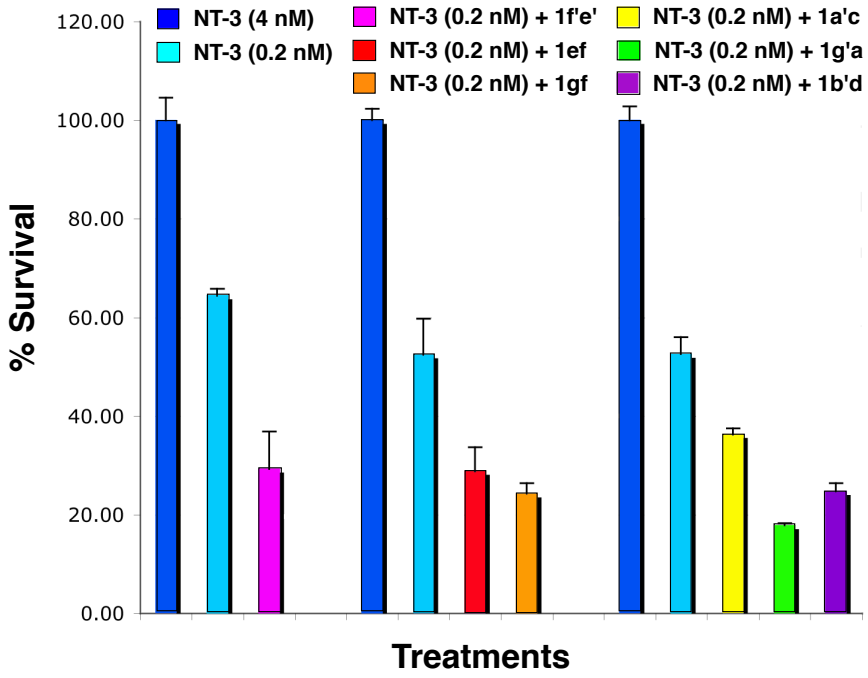
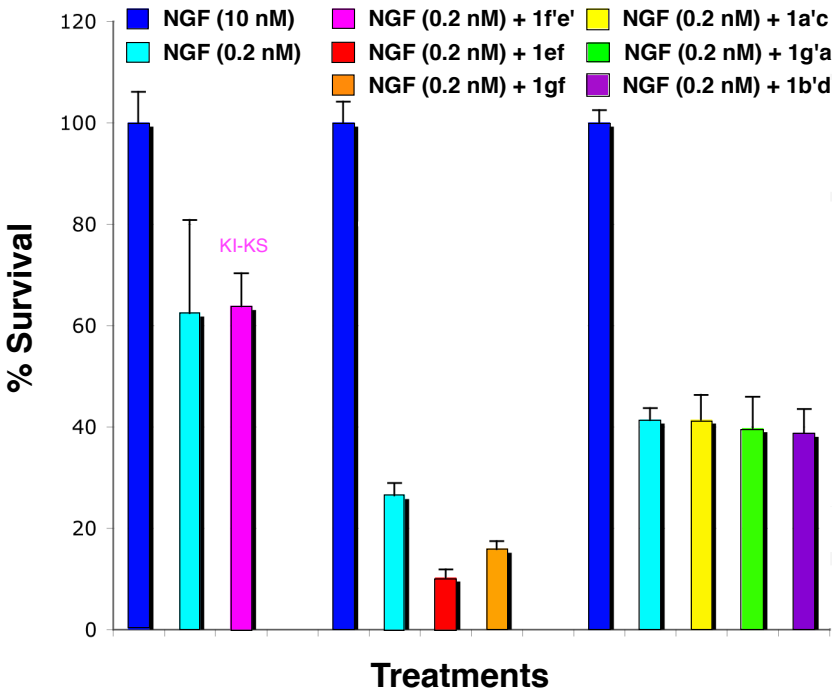


Figure 2.8. Selective antagonism of cell survival (a) NIH-3T3 cells expressing TrkC were cultured in SFM supplemented with the indicated peptidomimetic (10 μM) with suboptimal (0.2 nM) or optimal (4 nM) concentrations of NT-3; and, (b) NIH-3T3 cells expressing TrkA were cultured in SFM supplemented with the indicated peptidomimetic (10 μM) with suboptimal (0.2 nM) or optimal (10 nM) concentrations of NGF; and, (c) neuronal nnr5-TrkC cells expressing TrkC and p75 and 4-3.6 cells expressing TrkA and p75 were cultured in SFM supplemented with the indicated peptidomimetic (10 μM) with suboptimal (0.2 nM) concentrations of NGF or optimal (4 nM) concentrations of NT-3.

b



c

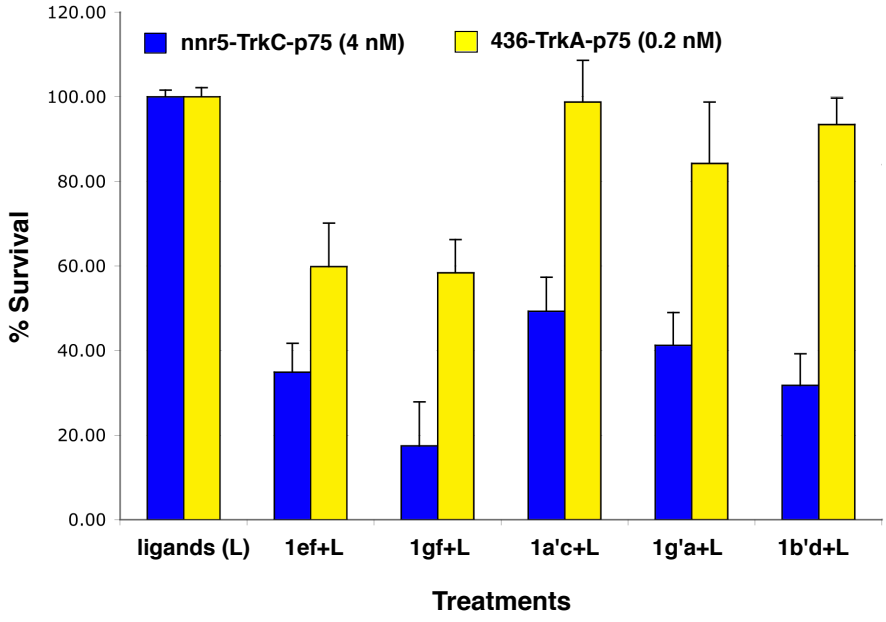


Figure 2.8. continued.

2.4.2 Signal Transduction Assays

The six active compounds **1ef**, **1gf**, **1f'e'**, **1a'c**, **1g'a**, and **1b'd** from survival assay were tested in signal transduction assays to confirm the antagonistic effects for TrkA or TrkC. The tyrosine phosphorylation of the receptors, phosphorylations of Akt, and MAPK were studied in Western blots of cell lysates prepared after stimulation of cells with the appropriate growth factor with or without compounds (Figure 2.9). In NIH-TrkC cells, NT-3 induces strongly TrkC-pTyr, and the phosphorylation of Akt and MAPK. **1ef**, **1gf**, **1f'e'**, **1a'c**, **1g'a**, and **1b'd** decrease these signals in both NIH-TrkC cells, and nnr5TrkC cells. Compounds **1ef**, and **1gf** also reduced NGF signals in TrkA-pTyr. None of the compounds affected the activating signals of IGF-1.

The biochemical data indicate that the peptidomimetics can inhibit NT-3 or NGF-dependent activation, tyrosine phosphorylation of TrkC or TrkA, and downstream signals of the receptors.

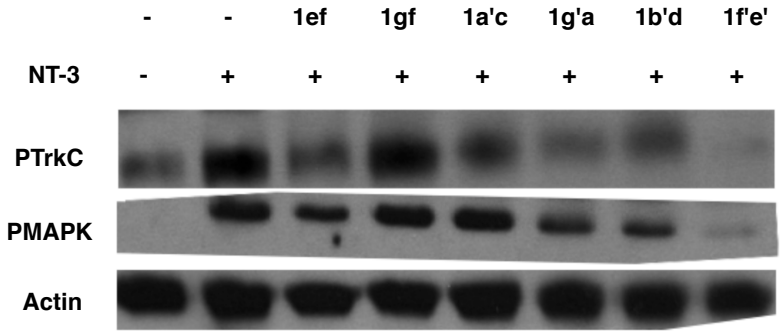


Figure 2.9. Inhibition of TrkC and TrkA receptor phosphorylation and signaling pathways. NIH-TrkC cells were exposed to the indicated peptidomimetics (10 μM) and NT-3 (0.2 nM) for 20 min. Detergent lysates were analyzed by Western blotting with anti-pTyr mAb 4G10 or anti-phospho-MAPK or anti-phospho-Akt. After stripping, membrane was reprobed with anti-actin to standardize loading. Blots were quantified by densitometry.

2.4.3 Binding Assays with Biotin-labeled Mimics

For binding assays, **1ef**, **1gf**, **1f'e'**, **1a'c**, **1g'a**, and **1b'd** were tagged with biotin instead of TEG. FACS binding assays indicated that the six compounds bind preferentially to the TrkC receptor. Compounds **1ef** and **1gf** were shown to be antagonist for TrkA and C in the cell survival assays described above; however, in the FACS binding data shown in Figure 2.10 does not show that these compounds bind to the TrkC expressing cells with much more affinity than other compounds **1f'e'**, **1a'c**, **1g'a**, and **1b'd**.

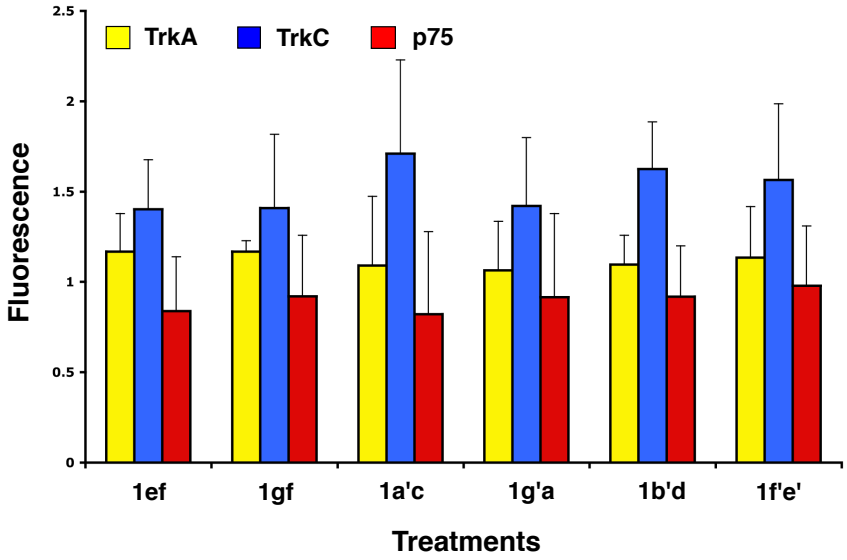


Figure 2.10. FACS binding of selected biotinylated peptidomimetics. Cell expressing the indicted receptors were first bound at 4 ^{o}C with the test ligand (20 μM), followed by fluorescein-avidin. After washing, cells were analyzed by FACScan/CellQuest. The background MCFs of NIH-IGF-1R were substracted to analyze the specific MCF binding to test cells.

2.5 Conclusion

In previous work, triazole-based β-turn bivalent mimics showed that four compounds selectively bind to TrkA. To improve the binding affinity and selectivity, we attached long alkyl chain linkers between triazole core and triazine core, focused the side-chains on *i*+1 and *i*+2 sequences in β-turn regions of neurotrophins, and gave a variation at the sequence orientation.

The 120 bivalent peptidomimetics were synthesized, and tested for overexpressed TrkA, TrkC, and IGF-1R cells. **1f'e'**, **1a'c**, **1g'a**, and **1b'd** showed a selective antagonism for TrkC. **1ef**, and **1gf** act as antagonists for both TrkA and TrkC. These compounds bind to the receptors, TrkA or TrkC, but not IGF-1R. The biology data showed the compounds are competitive antagonists for NGF or NT-3. That is, the compounds block the binding of natural ligand to the receptors. The fact that the compounds do not reduce the receptor-mediated activities without NGF or NT-3, and they do not affect neuritogenic differentiation, mean the compounds act as partial antagonists for the receptors, and the antagonisms are limited to ligand-depentant cell survival.

In previous miminalist mimics, four bivalent mimics selectively bind to TrkA. However, in the new library, four compounds **1f'e'**, **1a'c**, **1g'a**, and **1b'd** selectively bind to TrkC, and two compounds **1ef**, and **1gf** bind to both TrkC and TrkA. We hypothesized that amino acid side-chains corresponding to β-turn in neurotrophins might give selectivity of the compounds for Trk receptors. Comparison between sequences of active compounds and neurotrophins were summarized in Table 2.3. **1gf** that binds to both TrkA and TrkC has dipeptide sequences corresponding to NGF and NT-3 turn regions, but **1ef** has a dipeptide that corresponds to turns in NT-3 but does not have ones found in the NGF. **1a'c** and **1g'a** that selectively bind to TrkC have sequences corresponding to NT-3, but **1f'e** and **1b'd** that also bind to TrkC selectively have sequences for in turns of NGF and BDNF. Consequently it is hard to explain the selectivities of these compounds. However, one observation may be relevant for this; Table 2.3 shows bivalent mimics with *N*-to-*C* orientations that correspond to *i*+1-*i*+2 in

NGF and NT-3 bind to TrkA and TrkC, but bivalent mimics with the opposite, *C*- to *N*-, allignment selectively bind to TrkC.

Table 2.3. Comparison between Sequences of Antagonists and Neurotrophins.[a]

compound	**1ef**	**1gf**	**1f'e'**	**1a'c**	**1g'a**	**1b'd**
	TrkA and C antagonists		TrkC antagonists			
R^1R^2-R^3R^4	SK-IK	IR-IK	KI-KS	GT-EK	RI-TG	SM-GK

[a] Red color in sequences corresponds to sequences in NGF, pink corresponds to BDNF, and blue corresponds to NT-3.

Overall, these results show that the changes of linker length, sequences, and sequences orientation in small molecules can give changes of binding selectivity for Trk receptors. In the next project, we want to develop application of the compounds into the selective binding for target cells.

CHAPTER III

TRIAZOLE-BASED MIMICS WITH CYTOTOXIC COMPOUND AS TUMOR TARGETING LIGANDS

3.1 Introduction

All the FDA-approved drugs for chemotherapy for neuroblastoma, pancreatic cancer, and non-small cell lung carcinoma are poisons (Figure 3.1). These cytotoxic drugs have therapeutic indices that favor cancer cell death over destruction of healthy tissue, but, nevertheless, they are very toxic substances. It is alarming that development of small molecule chemotherapies has never broken out of this particular paradigm.

Existing "smart", targeted cytotoxic agents mostly feature antibodies directed to cell surface receptors.[63,64] Such protein-based targeting agents are expensive to produce on-scale, and are vulnerable to proteolytic degradation *in vivo*. Consequently, it is highly desirable to substitute antibodies with conveniently accessible small molecules. Like antibody conjugates, these agents could be injected. Many small molecules that bind cell-surface receptors can be imported into cells as the activated receptor internalizes, perhaps more easily than antibodies can. Finally, unlike large proteins, most small molecules are not degraded by proteases and are not immunogenic.

Despite the attributes of organic targeting agents, there are relatively few for over-expressed tumor cell surface receptors; this limits design and innovation. A reason for this situation is that there is no approach to facilitate discovery and production of small molecule non-peptidic targeting probes that parallels the convenience and reproducibility of antibody production in animals.

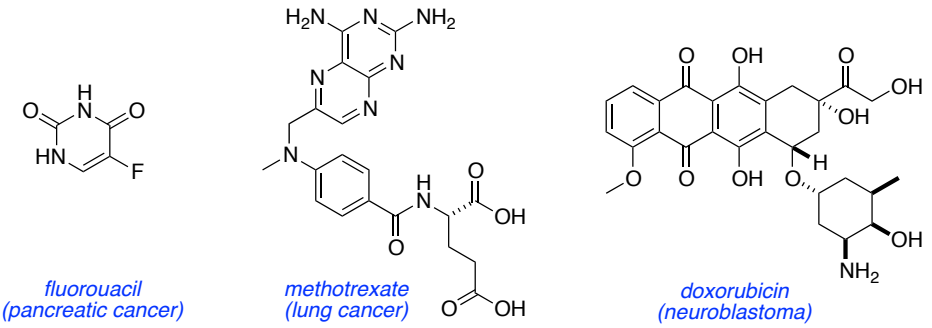


Figure 3.1. Examples of FAD-approved drugs for pancreatic cancer, lung cancer, and neuroblastoma.

Our group designed triazole-based bivalent mimics that target the TrkC receptor.[48] Compounds KB1341-1349 in Figure 3.2 selectively bind overexpressed TrkC cells; some showed partial agonist effects for TrkC, and others induced neurogenic differentiation. TrkC receptors are overexpressed in cancers such as prostate, medulloblastoma, neuroblastoma, malignant melanoma, and pancreatic cancer.[39] Thus, we thought KB1341-1349 and related materials could deliver cytotoxic compounds or imaging probes to cancers expressing the TrkC receptor.

KB1341-1349 (corresponding to these resideus, R^1R^2-R^3R^4: IY-LW, IY-LW, IY-IY) were conjugated with cytotoxic agents (specifically, 6-mercaptopurine and a rosamine) via click reactions, then their relative cell-killing effects were tested for both NIH-3T3 cells stably transfected with the TrkC receptor ($TrkC^+$ cells), and the corresponding wild-type ($TrkC^-$ cells or WT). These experiments were designed to test for selective delivery to $TrkC^+$ cells but not to $TrkC^-$

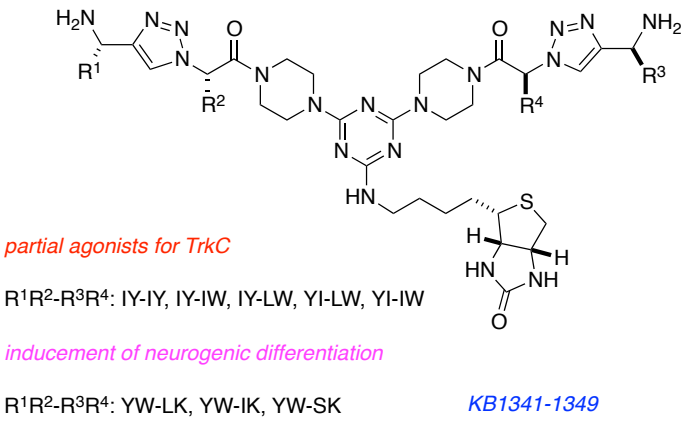
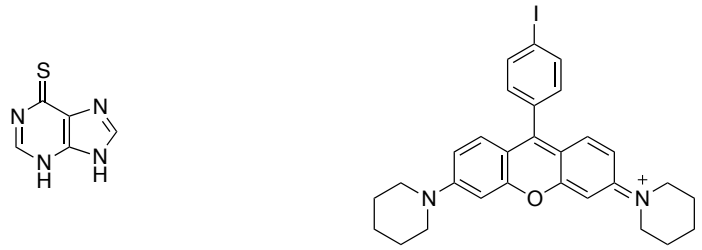


Figure 3.2. Bivalent mimics designed to selectively target the TrkC receptor.

6-Mercaptopurine (6-MP; Figure 3.3) has been used as an anti-cancer drug to treat leukemia, pediatric non-Hodgkin's lymphoma, and inflammatory bowel disease.[65-67] It inhibits purine nucleotide synthesis and metabolism. In our work, *N*-9 of the compound was alkylated to enable it to be connected (via an ethylene glycol linker, in fact), and to add azide functionality for the click reaction.

Our group reported rosamine analogs that have anti-proliferatic activity against leukemia and solid tumors.[68] One of these (Figure 3.3) showed an IC_{50} of 0.09 μM for the promyelocytic leukemia cell line, HL-60. This compound is very cytotoxic, water soluble, and it can be used for cell imaging. To do this, we changed the iodine group to azide facilitating click reactions.

The monovalent mimics (scaffolds with two amino acid side chains) used to make **4** were synthesized by Dr Dianjun Chen in our laboratory.[48] Rosamine azide **6** were prepared by Dr Liangxing Wu[69], and Anyanee Kamkaew, and BODIPY azide **8** was prepared by Anyanee Kamkaew. TrkC overexpressed NIH-3T3 cells were kindly provided by Dr David Kaplan at University of Toronto in Canada, and wild-type NIH-3T3 cells provided by Dr Jean-Philippe Pellois at Texas A&M University.



6-mercaptopurine *rosamine analog*

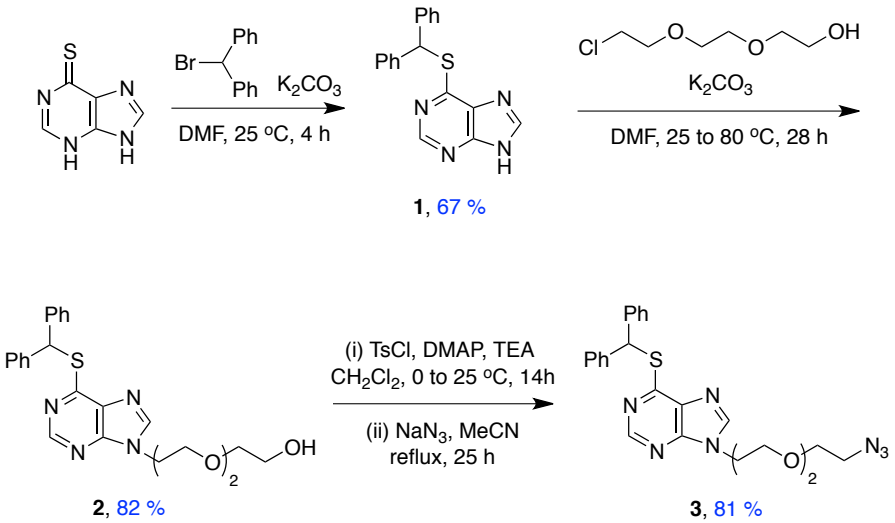
Figure 3.3. 6-Mercaptopurine (6-MP) and rosamine analog (Ros).

3.2 Syntheses and Biological Data of Bivalent Mimics with 6-Mercaptopurine

The 6-MP containing bivalent mimics **5** were "clicked" with an alkyne-modified TEG tag as previously described by our group.[47] Protection of the thiocarbonyl group of 6-MP with bromodiphenylmethane was used to obtain compound **1**.[70] 2-[2-(2-Chloroethyoxy)-ethoxy]ethanol was added, then modified to an azide via tosylation then treatment with sodium azide (Scheme 3.1a). Click functionalization using $CuSO_4$ and sodium ascorbate at 25 ^{o}C did not give the desired product **5**. Consequently, we increased the reaction temperature to 50 ^{o}C, and, in a separate experiment, added catalytic TBTA at 25 °C; both these sets of conditions worked well. Purification of the crude products from these reactions was achieved via RP-preparative HPLC. Deprotection of diphenylmethyl group from thione was performed with 50% TFA/CH_2Cl_2 and 3 equimolar phenol at 25 ^{o}C, and then the mixture was purified with RP-preparative HPLC. The final products **5** were lyophilized three times in 1.0 % acetic acid to remove TFA. Table 3.1 shows a summary of side-chain structures of bivalent mimics with 6-mercaptopurine.

Scheme 3.1. Preparation of Bivalent Mimics containing 6-Mercaptopurine **5**.

a

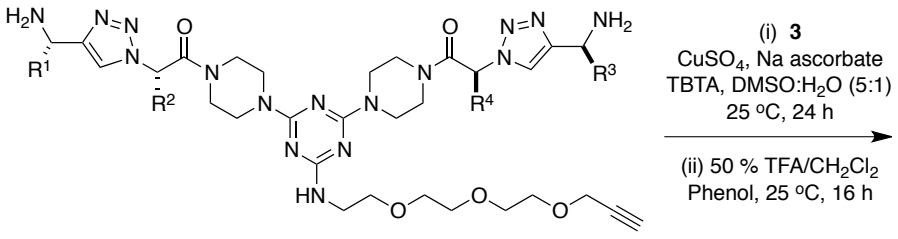


1, 67 %

2, 82 %

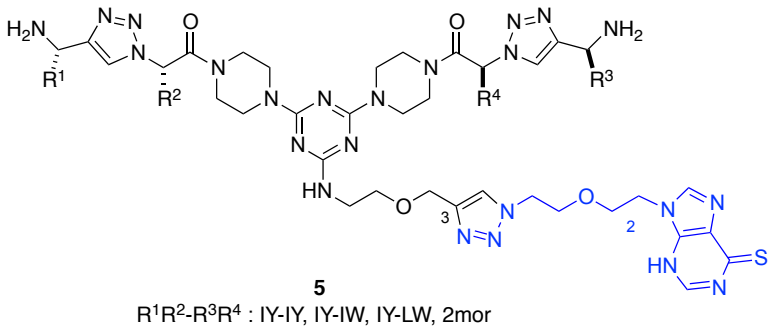
3, 81 %

b



4

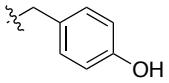
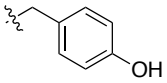
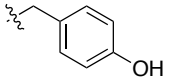
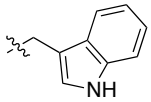
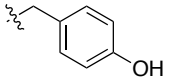
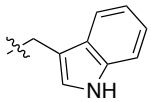
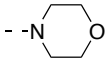
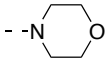
R^1R^2-R^3R^4 : IY-IY, IY-IW, IY-LW, 2mol



5

R^1R^2-R^3R^4 : IY-IY, IY-IW, IY-LW, 2mor

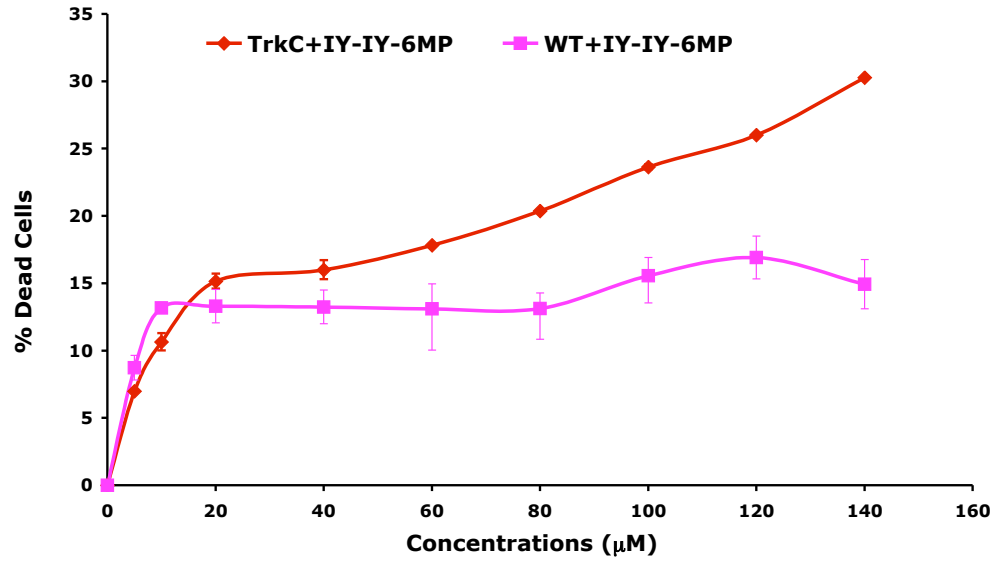
Table 3.1. Side-chain Structures of Bivalent Mimics with 6-Mercaptopurine **5**.

com'ds	sequence	R^1	R^2	R^3	R^4
a	IY-IY				
b	IY-IW				
c	IY-LW				
d	2mor				

Four bivalent mimics with 6-MP (**5a-d**) and four with TEG tags (**4a-d**) were tested to investigate targeted cytotoxicity. TEG labelled compounds **4a-d** were used as controls lacking the toxic part, 6-MP, and "morpholine capped" agents **5d** and **4d** were used to control for peptide parts. An antiproliferative assay was used to test **5a-d** on $TrkC^+$ cells (TrkC) and $TrkC^-$ cells (WT; Figure 3.4a). It emerged that **5a** is more cytotoxic in $TrkC^+$ cells than $TrkC^-$ cells. Further evidence for targeted delivery of **5a** was obtained via comparisons with the control compounds mentioned above (Figure 3.4b); these controls did not show meaningful cytotoxicities for TrkC cells or WT cells. Water solubility issues prevented determination of IC_{50} values for **5a**, **5b** and **5d** (for instance, at concentrations >140 μM **5a** precipitated in the medium).

It is known that the cytotoxicity of 6-MP is reduced when it is alkylated in the *N*9 position, as in **5a-d**.[65,71] For $TrkC^+$ cells, we determined 6-MP was has a cytotoxicity of 10 - 20 μM, but derivatives with an alkyl group to connect the ligands had cytotoxicities that were much less; thus **5a-d** had cytotoxicities that are significantly less than parent 6-MP; this was the main reason we decided to use another cytotoxic compound, the rosamine mentioned above.

a



b

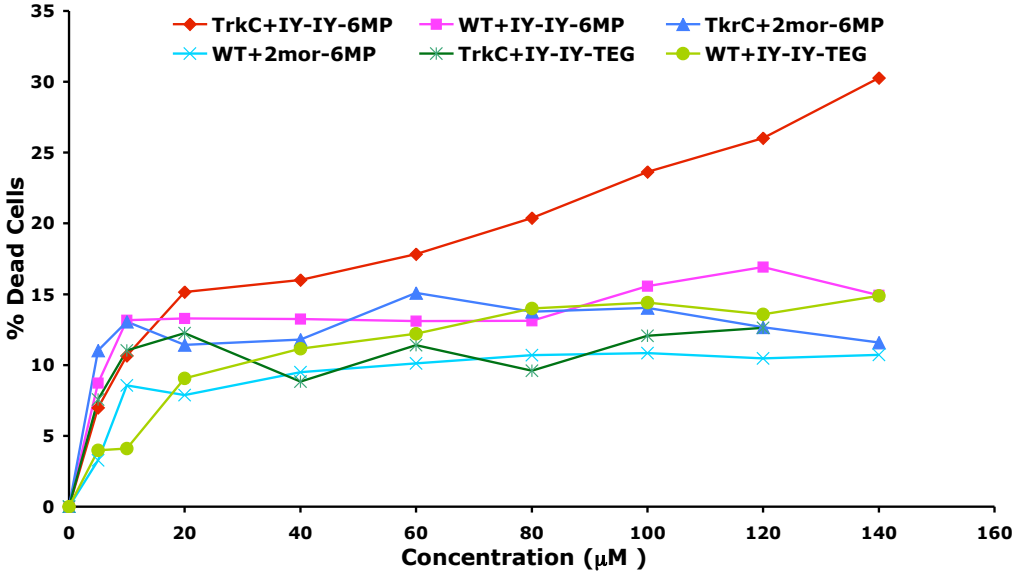


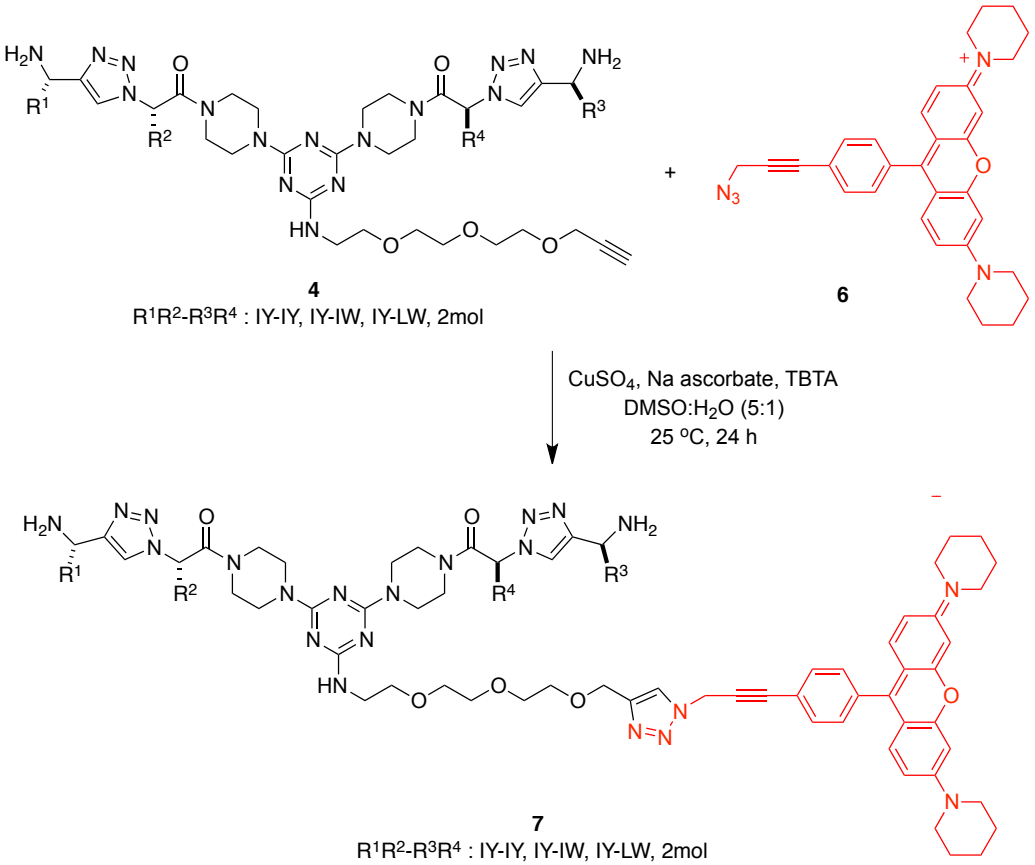
Figure 3.4. (a) Antiproliferative assay comparison for the compound with mercaptopurine **5a** (IY-IY-6MP) in MTT assays featuring TrkC-overexpressed NIH-3T3 cells (TrkC) and wild-type NIH-3T3 cells (WT). (b) Antiproliferative assay comparison for **5a** (IY-IY-6MP), **5d** (2mor-6MP), and **4a** (IY-IY-TEG) in MTT assays using TrkC and WT.

3.3 Syntheses and Biological Data of Bivalent Mimics with Rosamine, and BODIPY

Previous work from these labs led to fluorescent, water-soluble rosamine derivatives with high cytotoxicities.[68] Thus, we used one of these rosamines in place of 6-MP to give fluorescent, water-soluble analogs. Conjugations were performed for the rosamine **6** with **4** (Scheme 3.2a; conditions as for **5a - d**), and for BODIPY to **4a** to obtain **9a**. That non-cytotoxic BODIPY was used for intracellular imaging; its solubility in the desired "click-solvent", 5:1 DMSO:H_2O, was low so only DMSO was used.

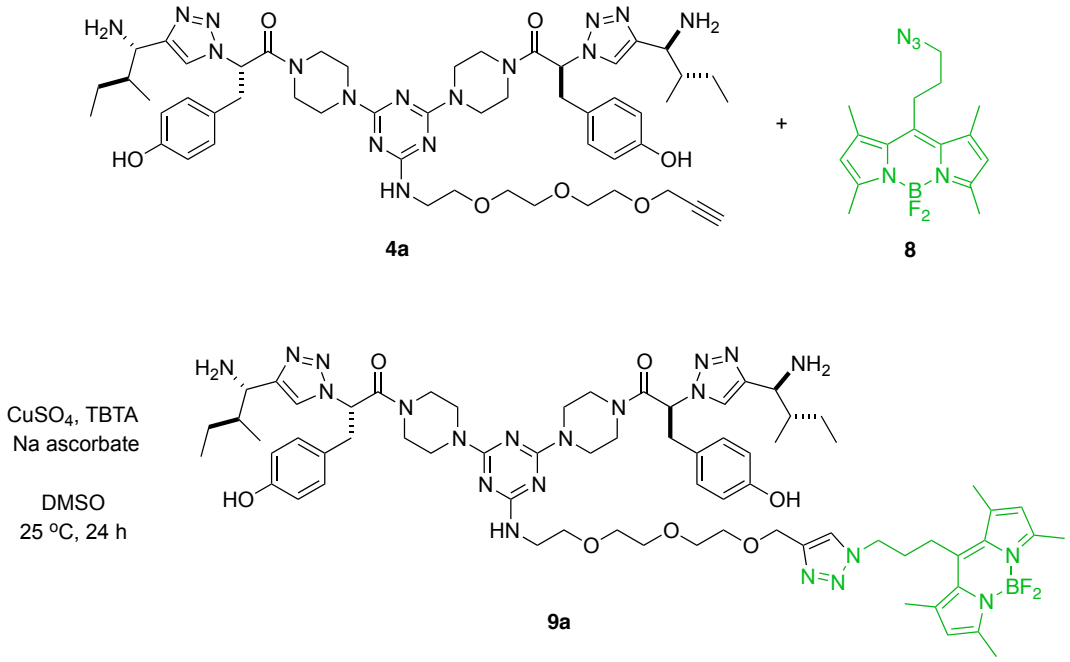
Scheme 3.2. Preparation of Rosamine Bivalent Mimics **7a - d** and The BODIPY One **9a**.

a



Scheme 3.2. continued.

b



Biological assays and cell imaging experiments for this work were conducted by Anyanee Kamkaew in our labs. All the series **7a - d** were tested in an antiproliferative assay on $TrkC^+$ cells (TrkC) and $TrkC^-$ cells (WT). Figure 3.5 shows only the data for **7a** on the two key cell types, but Table 3.2 shows a summary of IC_{50} values for all compounds **7a - d**. The best result came from compound **7a**; (IC_{50} 15.80 μM for $TkrC^+$ cells, and 27.58 μM for $TrkC^-$ cells). Overall the rosamine labeled compounds **7a - c** were shown to be more cytotoxic for $TrkC^+$ cells than for $TrkC^-$ cells, although only small IC_{50} differences were observed for **7b** and **7c** on the two different cells. Significantly, the morpholine control compound **7d** did not show any cytotoxicity difference for two cells.

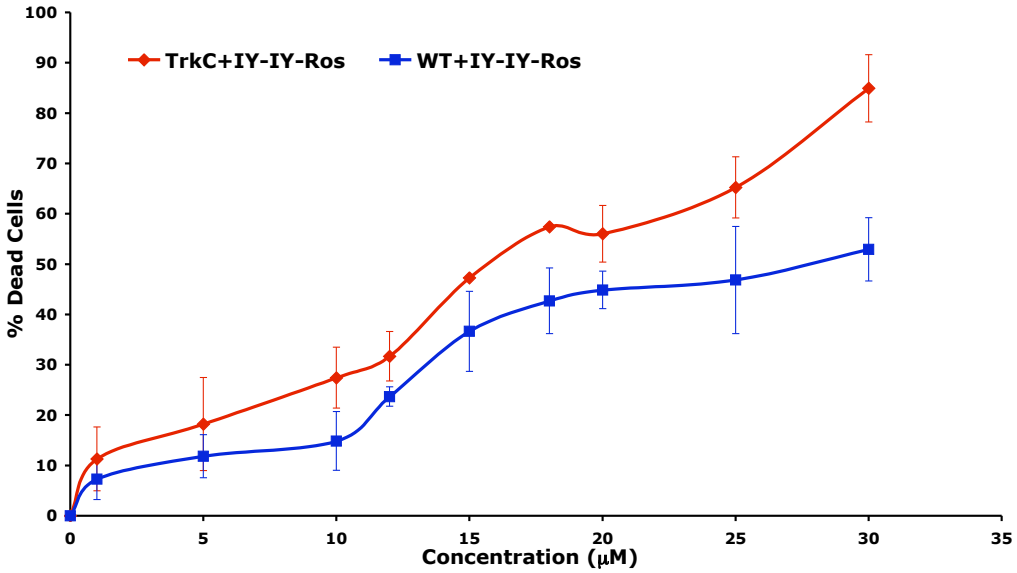


Figure 3.5. Antiproliferative assay comparison for the compound with mercaptopurine **7a** (IY-IY-Ros) in MTT assays featuring $TrkC^+$ cells (TrkC) and $TrkC^-$ cells (WT).

Table 3.2. IC_{50} Values Summary of Rosamine-labelled Compounds **7a-d** for $TrkC^+$ Cells and $TrkC^-$ Cells.

com'd	sequence	IC_{50} (μM) for $TrkC^+$ cells	IC_{50} (μM) for $TrkC^-$ cells
7a	IY-IY	15.80 ± 0.18	27.58 ± 1.38
7b	IY-IW	13.62 ± 1.09	19.00 ± 0.25
7c	IY-LW	14.08 ± 1.07	17.84 ± 0.94
7d	2mor	14.11 ± 0.71	14.99 ± 0.34

To prove that the rosamine-labeled compound binds to the TrkC receptor, we tried to do competitive cytotoxicity test with NT-3 (known natural ligand for TrkC

receptor), and IY-IY-TEG **4a** (a partial agonist for TrkC receptor)[48]. This assay used 3.5 nM of NT-3, and 20 μM of **4a**; the concentrations are sub-optimal doses that are obtained from cell-survival assays (data not shown). In TrkC cells, cytotoxicities by **7a** were reduced under existence of NT-3 and **4a** (Figure 3.6a). However, the ligands did not affect cytotoxicities by **7a** in WT cells (Figure 3.6b). That is, **7a** competed with NT-3 and **4a** in TrkC cell, but not in WT cells.

a

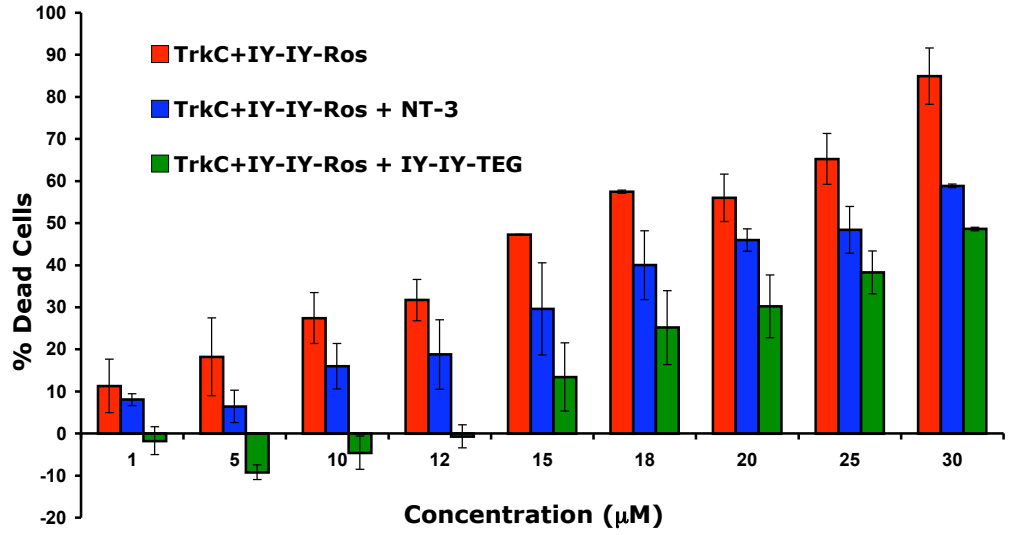


Figure 3.6. (a) Competitive cytotoxicity of **7a** (IY-IY-Ros), **7a** with NT-3 (IY-IY-Ros+NT-3), and **7a** with **4a** (IY-IY-Ros+IY-IY-TEG) in TrkC cells; and, (b) in WT cells.

b

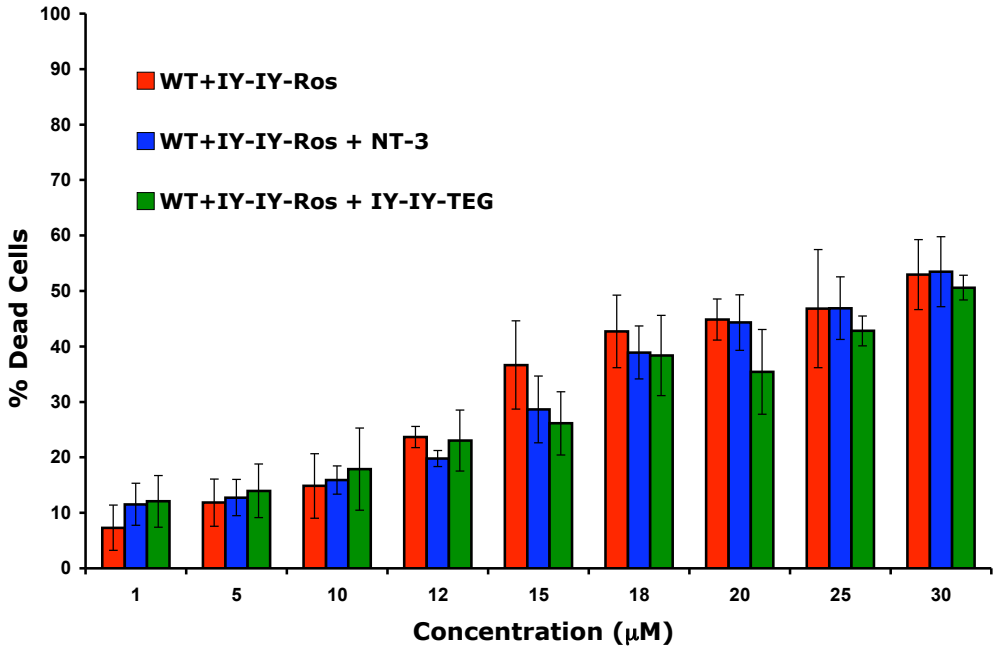
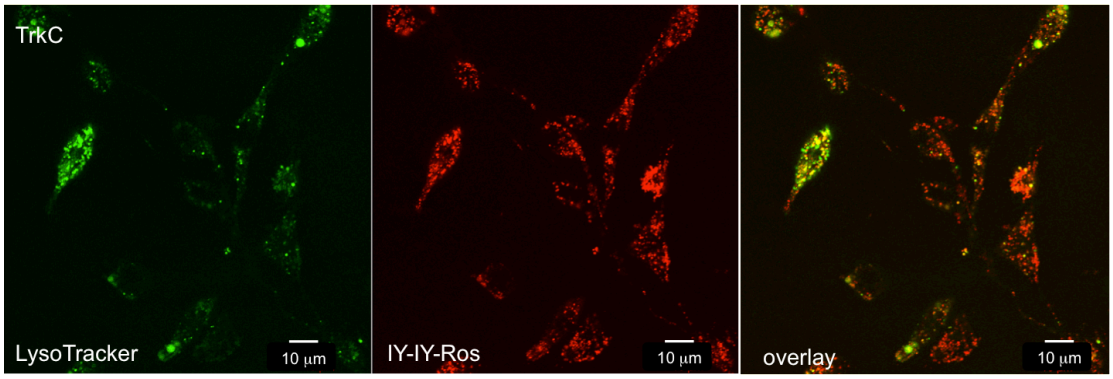


Figure 3.6. continued.

Next, we took fluorescence from rosamine or BODIPY to investigate if the compounds with dyes are internalized into the cells and if so, where the compounds are located. Previous work reported rosamine anolog itself is localized in the mitochondria.[68] However, rosamine-labeled compounds **7a** - **c** were localized in the lysosome, but not the mitochondria (Figure 3.7a). We confirmed this result with BODIPY-labeled compound **9a**, because co-localization of rosamine-labeled compounds with Lyso-Tracker gave weak signals (Figure 3.7a and 3.7b). On the other hand, "morpholine cap" compound **7d**, a negative control without any peptide, accumulated in the mitochondria like rosamine analog alone. The results correspond to Bartheld's report about subcellular pathways of NT-3 that is a known ligand for TrkC.[72] Bartheld's group examed subcellular pathways and accumulation of internalized radiolabeled neurotrophins in retinal ganglion cells (RGCs). They reported two possible pathways of internalized NT-3; lysosome pathway and Golgi pathway. In the lysosome pathway, NT-3 that bound to TrkC is degraded in the lysosomes. Moreover, a research for

labelling densities (LD) in organelles of RGCs after intraocular injection of NT-3 in chick embryos showed LD in lysosomes (LD 7.73) was much higher than other organelles, and LD in mitochondria (LD 0.06) was very low compared to other organells. Our cell imaging data, in light of Bartherld's observation, allow us to conclude that our compounds **7a - c** follow the pathway of NT-3.

a



b

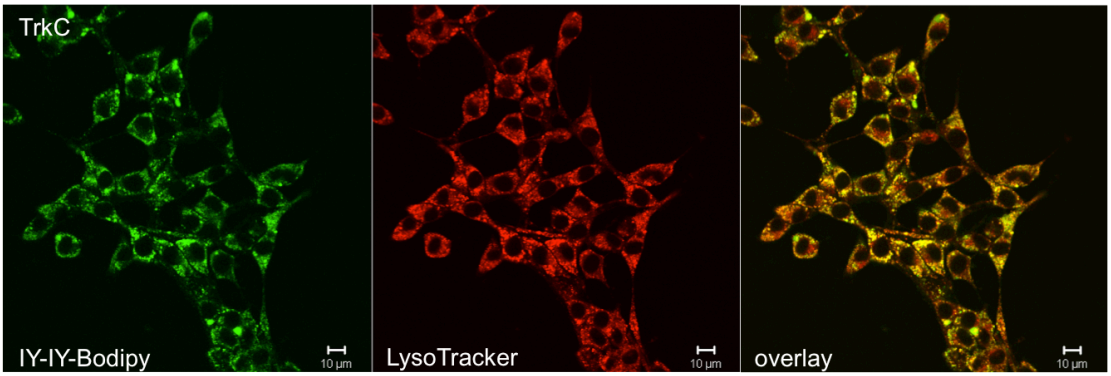


Figure 3.7. (a) Fluorescence of compound **7a** and Lyso-Tracker in TrkC cells; (b) compound **9a** and Lyso-Tracker; and, (c) compound **7d** and Mito-Tracker.

c

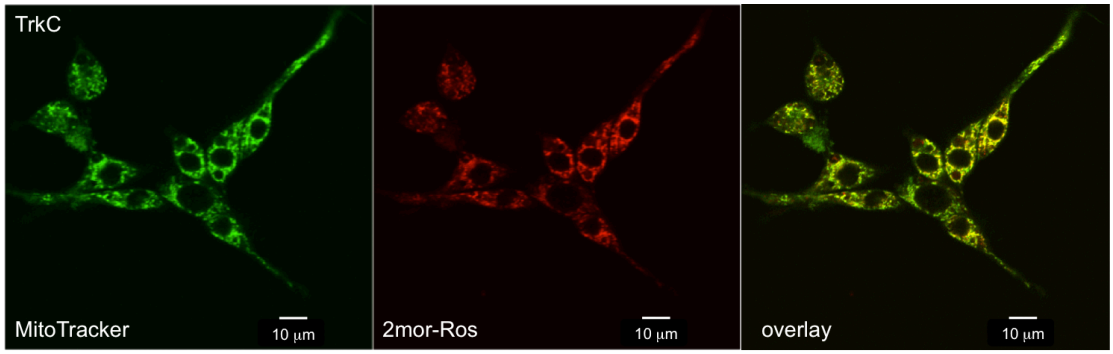


Figure 3.7. continued

3.4 Conclusion

It is important to develop accessable small molecules that selectively target cancer cells, because this would allow the small molecule to deliver anti-cancer drugs to the cancer effectively and decrease damage to healthy cells. With information about a type of cancer, such as expressed receptors or structure, targeting molecules can be designed. Previously, our group developed triazole-based bivalent peptidomimetics that mimic β-turn hot-spots of NT-3, and selectively bind to the TrkC receptor which is overexpressed at various cancers such as medulloblastoma, and neuroblastoma. To confirm the TrkC targeted delivery by the compounds, we synthesized **5a** - **d**, **7a** - **d**, and **9d** conjugated with cytotoxic compounds, 6-mercaptopurine (6-MP), and rosamine analog (Ros), and tested the antiproliferative assay for TrkC overexpressed NIH-3T3 cells, and wild-type NIH-3T3 cells. Overall data showed the bivalent peptidomimetics with cytotoxic compounds were more cytotoxic for TrkC cells than for WT cells. In addition, a competitive assay with NT-3 and IY-IY-TEG proved the bivalent peptidomimetics competed with the known ligands for TrkC. These results show that the triazole-based bivalent mimics deliver anti-cancer drugs to a target cancer cell selectively. We developed the targeting compound for TrkC overexpressed cells *in*

vitro, and, in next study, the compounds will be tested for targeting neuroblastoma which has overexpressed TrkC receptors in an *in vivo* assay.

In this study, we used information that some cancer cells have overexpressed TrkC receptors, to develop a targeting compound. However, rapidly proliferating cancer cells express many surface receptors that are not usually found on the surface of healthy cells. The functions and structures of most of these receptors are unknown. The next chapter describes the identification of small molecule ligands for unknown targeting cancer cells.

CHAPTER IV

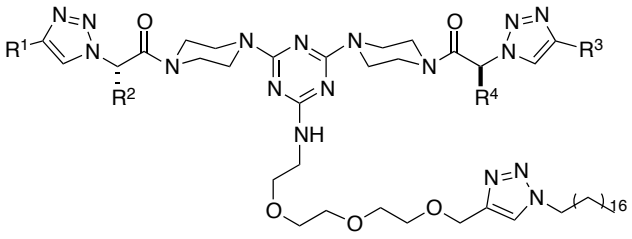
TRIAZOLE-BASED BIVALENT MIMICS THAT TARGET PANCREATIC CANCER, PREPARED WITH LUMINESCENT EUROPIUM(III) CHELATES TO QUANTITATE BINDING

4.1 Introduction

Cancer cells express a complicated set of cell surface receptors, and even the corresponding healthy cells have largely uncharacterized surface complexions. Thus, we regard each cell, healthy or tumor, as a living combinatorial library of receptors that may be targeted for small molecules.

Triazole-based bivalent mimics like Figure 4.1a are pairs of "monovalent" secondary structure units each expressing two side-chains, as described in Chapter 2. In a previous study from our group, 15 homodimeric and 135 heterodimeric bivalent mimics were prepared for "The Luciferase Assay" (Figure 4.1b).[73] Liposomes bearing the compounds were expected to target undetermined cell surface features in tumor cells. Each compound had a hydrocarbon tail to allow them to be anchored into liposomes with an entrapped luciferase reporter-gene-cargo, simply by mixing and incubation overnight; this was done in a one-per-well format. Incubation of the functionalized liposomes with cells resulted in a background rate of delivery of the luciferase gene, and bioluminescence (on lysis and treatment with appropriate reagents). If the bivalent ligand expressed on the surface of the liposome targets a cell surface receptor, then more cargo is delivered and bioluminescence is enhanced.

a



b

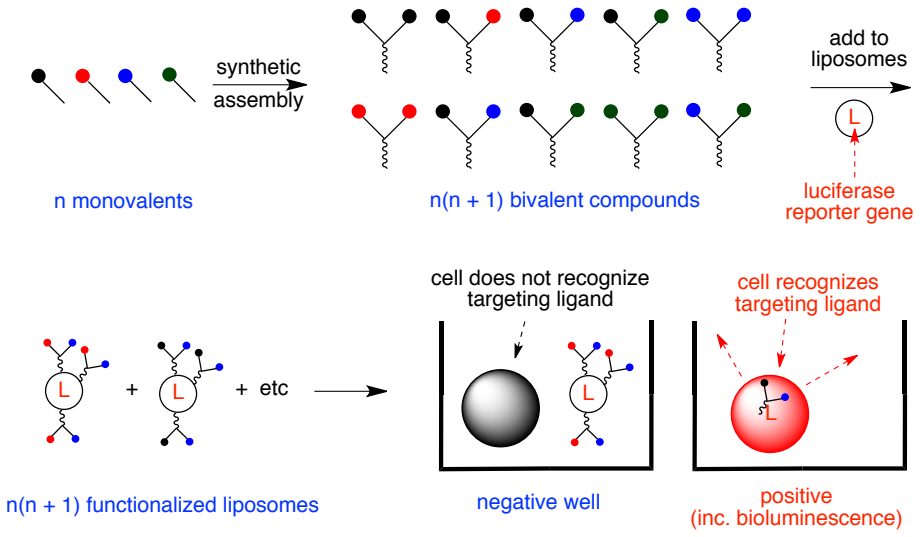
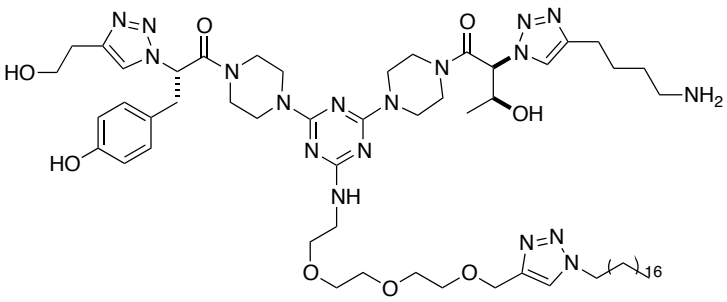


Figure 4.1. (a) Triazole-based bivalent mimics prepared for the luciferase assay. (b) The luciferase assay for identifying small molecules on liposomes that target tumor cells.

Tumor cells cultured on their own behave differently to ones "sub-cultured" in healthy cells (human umbilival vein endothelial cells, HUVEC) where the cell's receptors adapt to facilitate angiogenesis. Consequently, in this work, homogeneous tumor cells and those in subculture were screened to target both the tumors and the vasculature. Figure 4.2 shows that KB1005 targeted *homogeneous* PANC-1 cells, whereas KB1023 was shown to bind a *subculture* containing the pancreatic tumor cells PANC-1, and KB1061 targeted human H1299 non-small-cell lung carcinoma in

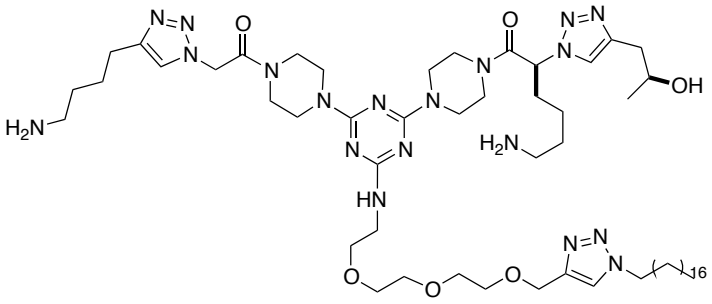
subculture. Neither KB1023 nor KB1061 targeted homogeneous PANC-1 or H1299 cells.

a



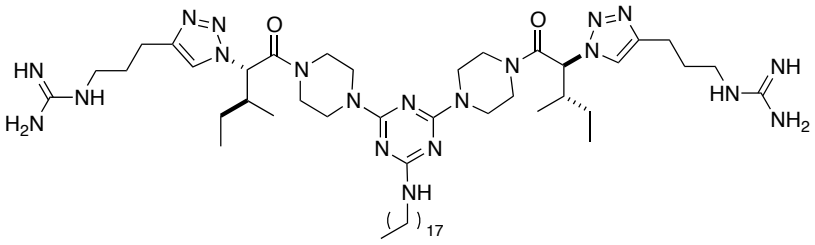
KB1023

b



KB1061

c



KB1005

Figure 4.2. (a) KB1023 targets PANC-1 subculture cells, but not PANC-1 alone. (b) KB1061 targets H1299 subculture cells, but not H1299 alone. (c) KB1005 targets only PANC-1 cells.

This chapter describes how a derivative of KB1005 was prepared to enable a homogeneous time resolved fluorescence (HTRF) assay to be used to determine a binding constant of the compound interacting with PANC-1 cells. Lanthanide-based assays are superior to ones based on radioactive isotopes, and normal fluorescent labels because the lanthanide lifetimes are longer than lifetimes of the background fluorescence from reagents or specimens (0.2 - 1.5 msec compared to ps - μs). Long lifetimes facilitate delayed measurements of emission signals at intervals when background fluorescence has completely decayed. Additionally, the large Stokes shifts of lanthanide complexes facilitate greater sensitivities, and their sharp emission peaks allow tight limits to be set on the excitation filters. For these reasons lanthanide-based fluorescent methods can have much higher sensitivities than those based on other dyes.[74,75]

Many groups have reported comparisons between traditional and lanthanide assays for ligand-receptor interactions.[76-79] For instance, Gillies compared a ^{125}I binding assay and the dissociation-enhanced lanthanide fluoroimmunoassay (DELFIA) for binding of NDP-α-MSH to the human melanocortin-4 receptor (hMC4R).[77] They synthesized Eu-DTPA-NDP-α-MSH (Figure 4.3) and compared its binding with that for the compound with ^{125}I-NDP-α-MSH, and NDP-α-MSH for hMC4R; the two competitive binding assays showed similar IC_{50} values (13 nM *vs* 23 nM), but the lanthanide assay showed higher sensitivity.

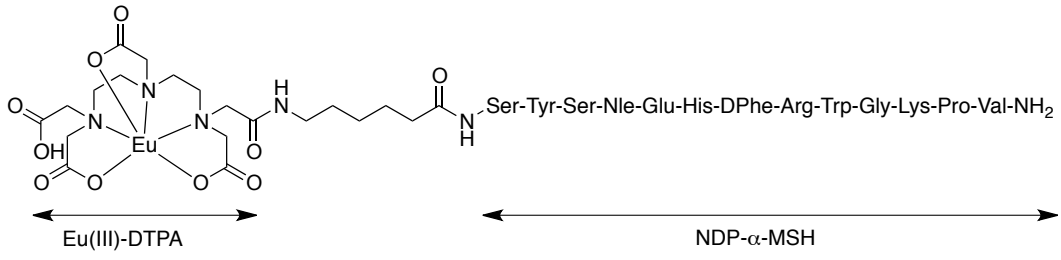


Figure 4.3. Eu-DTPA-NDP-α-MSH.

We followed Gillies' approach to establish HTRF for KB1005 modified with a Eu-chelate (Eu-PCTA, **1**; Figure 4.4). Simple TEG-labeled compounds were used for binding competition experiments. Four compounds, similar to KB1005 but with different side-chains, were synthesiged as controls to confirm the targeting for pancreatic cancer (in fact, they target other cancer cell types).

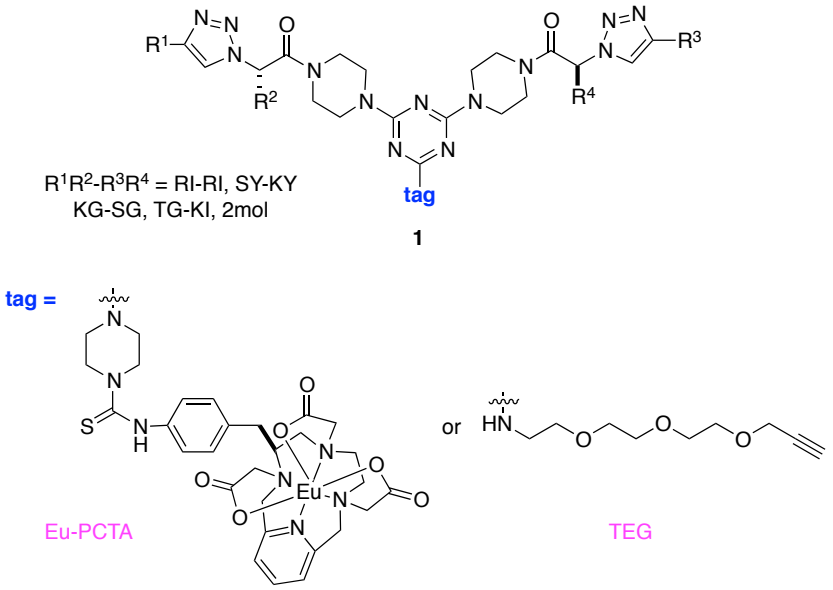


Figure 4.4. Bivalent mimics **1** with Eu(III)-chelate (Eu-PCTA) or TEG tag.

4.2 Syntheses of Bivalent Mimics with Luminescent Europium(III) Chelates

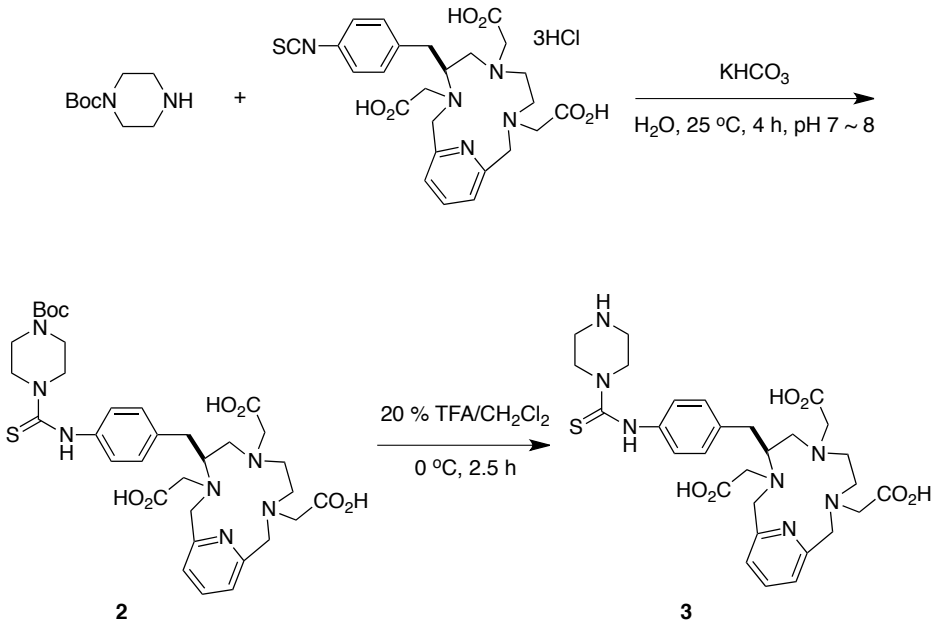
For the lanthanide-based assay, five bivalent peptidomimetics with Eu(III) chelates, and five bivalent peptidomimetics with TEG tag were synthesized. The TEG-labeled compounds were synthesized by the method described in Chapter 2.3. Monovalent mimics **4** were synthesized by Dr Yu Angell.[47]

Our group developed a method to conjugate monovalent mimics to bivalent mimics on triazine core as was described in Chapter 2. The general procedure is to

attach a tag first at cyanuric chloride via S_NAr reaction, and then add one monovalent mimic to the core at a time using solvent effects e.g. THF *vs* DMSO.[47,80] First, we tried the method illustrated in Scheme 4.1. Modified chelate **3** was successfully added on the core in the first step; all reactions were monitored by analytical HPLC, and MALDI-MS. However, product **7** could not be made under these conditions. Analytical HPLC showed the starting material **6** was consumed completely, but compound **7** did not form. Our initial thought was that several carboxylic groups in the chelate might be interfering with the S_NAr reaction. Thus, we tried to introduce Eu to chelate **2** first, and then to do the S_NAr reaction with a monovalent mimic **5**. However, we also failed to obtain **7**; product was a mixture of several compounds by HPLC, and MALDI-MS for the crude reaction mixture did not show anything corresponding to **7**.

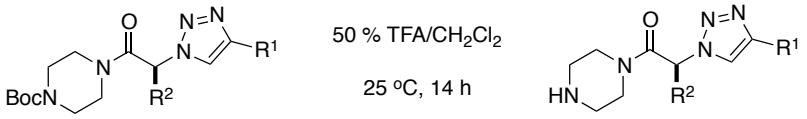
Scheme 4.1. First Attempted Syntheses of Preparing Eu-chelated Bivalent Mimics **1**.

a



Scheme 4.1. continued.

b

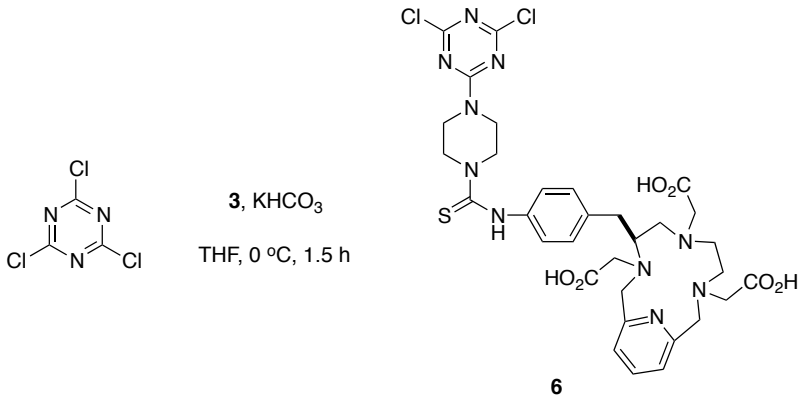


4

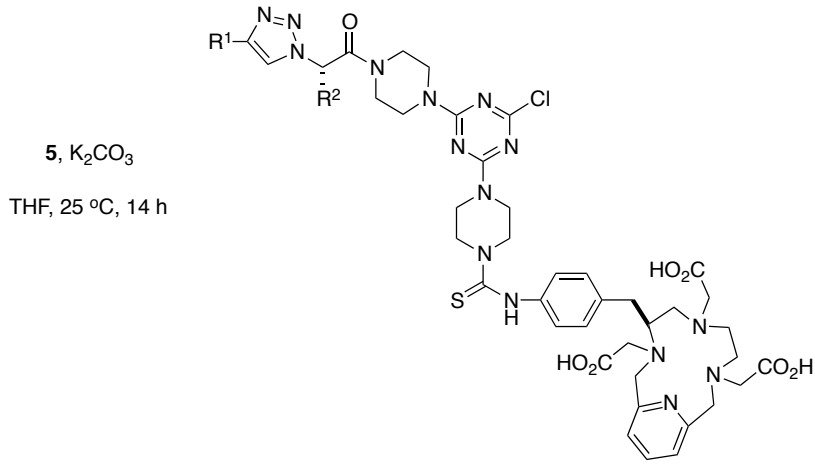
R^1R^2 = RI, SY, KT
KG, SG, TG, KI

5

c



6

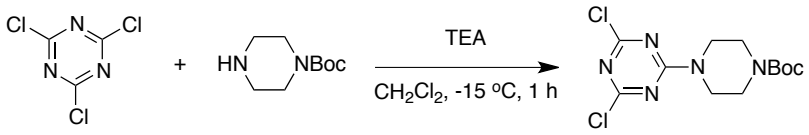


7

We believed that the chelate part was inhibiting the S_NAr reaction. Thus, we decided to add monovalent mimics first, and then the chelate in the last step. The synthesis was modified slightly, as shown in Scheme 4.2a. Boc-piperazine was added to cyanuric chloride first as a linker for the chelate, and then the bivalent mimics **9** were introduced via the previous method. Boc deprotection and addition of p-SCN-Bn-PCTA gave compound **10**. Finally, we obtained bivalent mimics with the chelate, PCTA, **10a** and **10e**. However, this method could not be used to make bivalent mimics with Lys side-chain **10b - d** (SY-KY, KG-SG, and TG-KI). These compounds after Boc-deprotection have two different amines, a secondary amine from piperazine and a primary amine from the Lys side-chain which can react with p-SCN-Bn-PCTA under the reaction condition. Thus, we modified the method again as shown in Scheme 4.2b; first, the monovalent mimics were added to cyanuric chloride via S_NAr reaction. In this case, it is also important that the monovalent mimic **3** without a Lys side-chain such as SY, SG, and TG is added to cyanuric chloride first. This is because the primary amine in Lys also reacts with cyanuric chloride via S_NAr reaction. Products **10b - d** were obtained by addition of the modified PCTA **3** to **12.** All compounds **10a - e** were purified with RP-preparative HPLC.

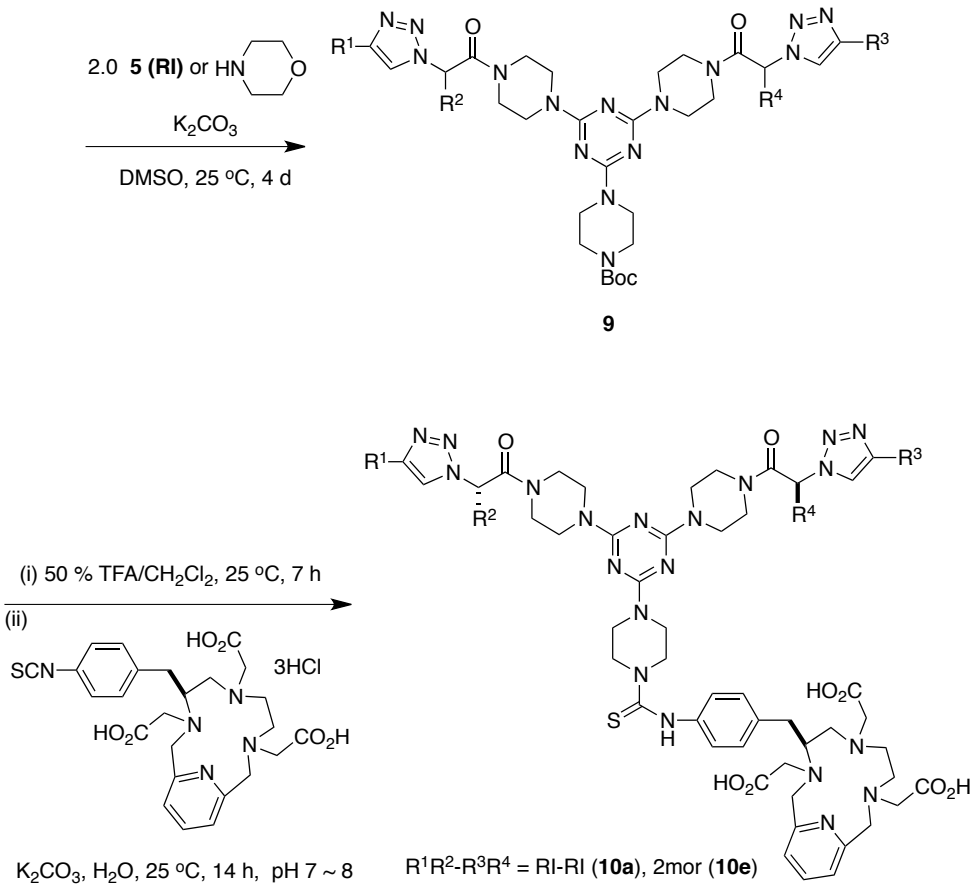
Scheme 4.2. Two Methods for Preparing The Bivalent Mimics **10a - e**.

a

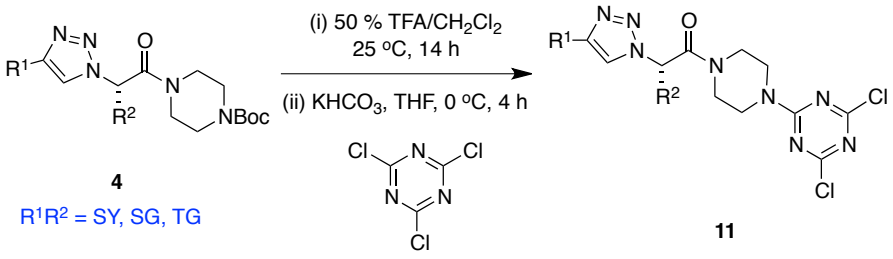


8

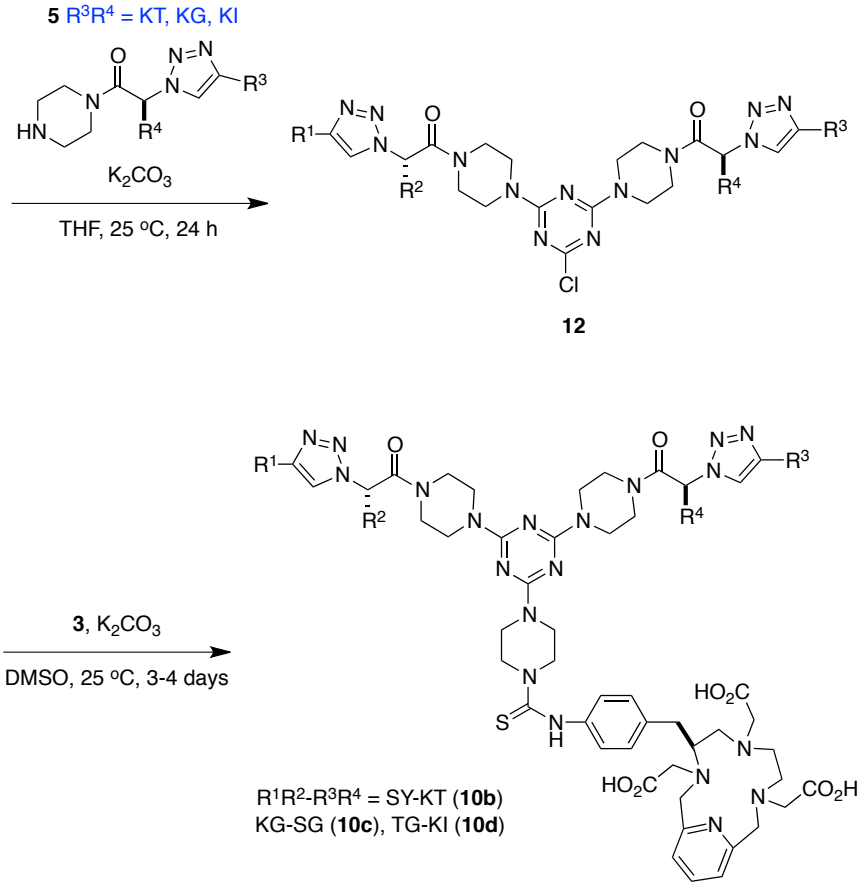
Scheme 4.2. continued.



b



Scheme 4.2. continued.



Europium was bound to compound **10** with an acetate buffer condition at pH 5.5. After the reaction, the product mixture was purified by RP-preparative HPLC, and compound **1** was obtained as a TFA salt. It is necessary to remove TFA for biology assays. Generally, the TFA salt form was lyophilized in 1.0 % acetic acid solution three times to replace the trifluoroacetate to acetate. However, analytical HPLC after using 1.0 % acetic acid showed some Eu escaped from the PCTA-chelate, reforming compound **10** under the acidic condition. Addition of Eu again after lyophilization gave one product, as indicated by analytical HPLC. However, we did not want to leave extra Eu in the compound. Thus, several other conditions were tried to solve this problem, such as 0.5 % or 0.1 % acetic acid, 50 nM ammonium acetate, 20 nM HCl and addition

of Eu again after lyophilization. Treatment of 1.0 % and 0.5 % acetic acid, 50 nM ammonium acetate, or 20 nM HCl gave split peaks after lyophilization. The best result was obtained by 0.1 % acetic acid. Therefore, lyophilization three times from 0.1 % acetic acid was used to remove the TFA. Table 4.1 shows a summary of side-chain structures of the prepared Eu-chelated bivalent mimics.

Scheme 4.3. Synthesis of Eu-labeled Bivalent Mimics **1a - e**.

a

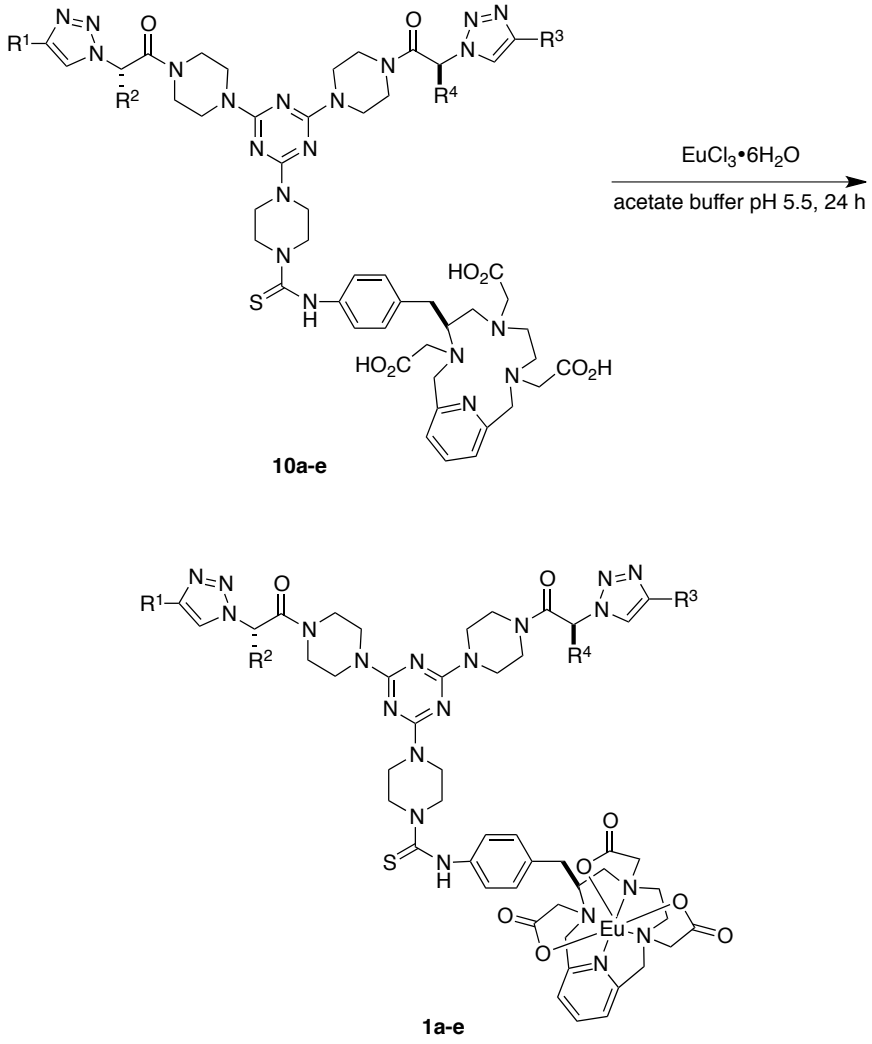
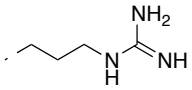
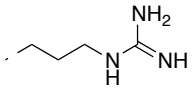
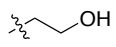
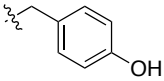
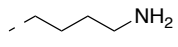
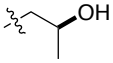
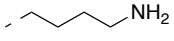
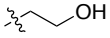
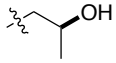
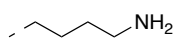
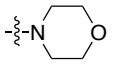
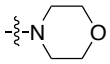


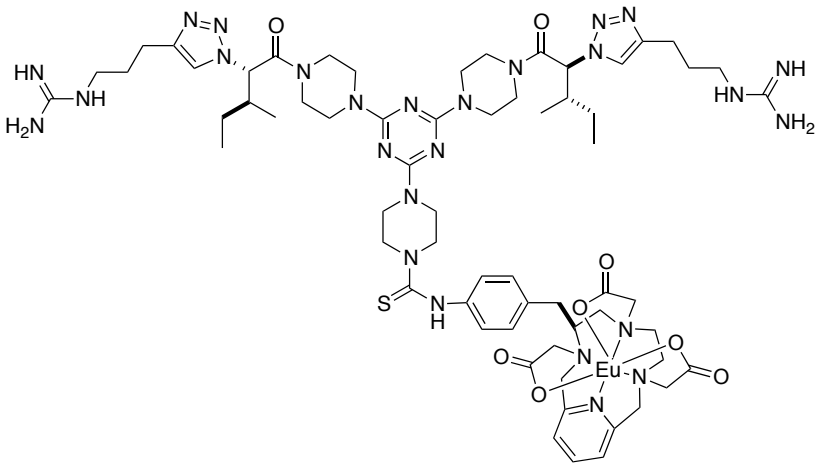
Table 4.1. Side-chain Structures of The Bivalent Mimics.

com'd	sequence	R^1	R^2	R^3	R^4
a	RI-RI				
b	SY-KT				
c	KG-SG		H		H
d	TG-KI		H		
e	2mor				

4.3 Biology Assay

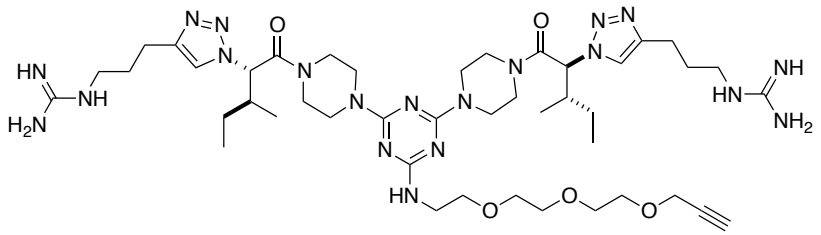
Total ten compounds including five Eu-labeled bivalent mimics and five TEG-labeled bivalent mimics were prepared for HTRF competitive binding assay. The assay was conducted by our collaborator, Dr Nancy Smyth Templeton and her coworkers at Gradalis Inc.. Figure 4.5 shows the structures of Eu-PCTA compound **1a** and TEG-labeled compound corresponding to KB1005 that targets pancreatic cancer selectively.

a



1a, *Eu-PCTA compound*

b



TEG-labeled compound

Figure 4.5. (a) Structure of Eu-PCTA compound **1a**; and, (b) TEG-labeled compound corresponding to KB1005 that targets pancreatic cancer.

The competitive binding assay used Eu-PCTA **1a-e** at a fixed concentration, 0.63 nM. The optimized concentration was determined by testing a range of 0.08 to 5.06 nM Eu-PCTA **1**. The unlabeled, non-fluorescent compound (TEG-labeled compound) was added to the wells of a 96-well dish in concentrations ranging from 1 to 50 nM to generate displacement curves. The binding constant was determined by using a nonlinear regression (Figure 4.6). The binding constant was K_a= 1.1 x 10^9 M^{-1}. Binding constants of 10^8 M^{-1} and greater are associated with small molecules that promote expression in the transfection-based assay. Ligands that meet the criteria for an

appropriate binding constant could be used for a personalized, targeted therapeutic in the clinic.

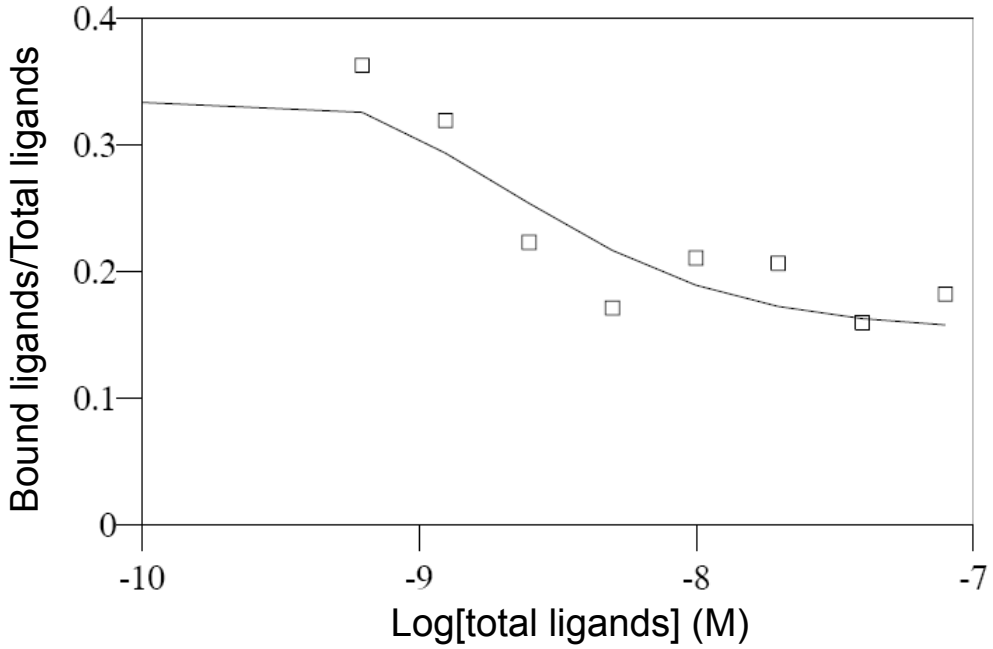


Figure 4.6. Competitive binding assay curve using Eu-PCTA compound **1a** and TEG-labeled compound.

4.4 Conclusion

It is important to develop small molecules that target only cancer cells, because the molecule can deliver anti-cancer drugs to the targeted cells effectively without damaging for healthy cells. In cases where we do not have much information about a target, a target can be found by screening with many compounds. In this strategy, our group found a small molecule KB1005 that targets pancreatic cancer, although we do not know how KB1005 interacts with the target yet. Next, we wanted to apply the

compound to advanced cancer environment, such as fresh surgical specimens of pancreatic cancer, and we wanted to establish a confirmatory homogeneous time resolved fluorescence assay (HTRF) to determine a binding constant for the cell. Thus, five Eu-labeled compounds **1a** - **e** and five TEG-labeled compounds were prepared for HTRF assay. The lanthanide-based assay was chosen to avoid background fluorescence from specimen. From the competitive binding assay, we found **1a** bound to the cell with $K_a = 1.1 \times 10^9\ M^{-1}$. This value can promote expression in the transfection-based assay.

Originally, the triazole-based peptidomimetics like KB1005 were designed as β-turn mimic for neurotrophins hot-spot regions, and showed some activities for TrkA in previous work.[47] We described these compound as "minimalist mimics", because the compounds mimic only side-chains of peptide secondary structures on a rigid core, e.g. triazole core, instead of a peptide back bone. At the same time, we proposed that sets of these types of compounds could mimic local pairs of amino acids in any secondary structures as "universal peptidomimetics". In our hypothesis, the compounds' library would be very useful for high-throughput assay against various targets because the compounds can mimic a range of secondary structures in known or unknown targets. As our hypothesis, KB1005 was found as an active compound for pancreatic cancer cells through high-throughput screening, and we followed up on this lead with more advanced studies. Next chapter describes the definition and validation of the minimalist mimics and universal peptidomimetics, and propose new scaffolds for them.

CHAPTER V

UNIVERSAL PEPTIDOMIMETICS*

5.1 Introduction

"Minimalist mimics" are compounds that mimic only side-chains of peptide secondary structures.[52] Scaffolds in the minimalist mimics do not exist as a single conformation in solution; they equilibrate between forms representing local minima in Boltsman distributions of energy states, so there is no reason that their global minimum should correspond to the target secondary structure. Instead, it is sufficient that the pertinent conformations for mimicry have energies similar to the global minima so that they are populated and the transition-state energy barriers to arrive at them can be overcome at ambient temperatures, that is, there are no insurmountable thermodynamic or kinetic obstacles to attaining the target conformations. Identification of conformations in equilibrating ensembles that are both kinetically and thermodynamically accessible can be done only by comparing with similar systems that have been studied experimentally or via computational method. Spectroscopic techniques like NMR cannot detect a preferred solution state conformer of mimic at room temperature, because there is none. Coupling constants and NOE measurements for the scaffolds would reflect conformational averaging. In addition, useful minimalist mimic cannot be totally flexible because their scaffolds must have limited degrees of freedom to avoid significant entropic penalites on adopting the target secondary structure conformations.

---

*Reprinted in part with permission from "Univeral Peptidomimetics", Eunhwa Ko, Jing Liu, Lisa M. Perez, Genliang Lu, Amber Schaefer, and Kevin Burgess, *J. Am. Chem. Soc.,* **2011**, *133*, 462-477. Copyright 2010 American Chemical Society. *Reprinted in part with permission from "Pyrrole-Based Scaffolds for Turn Mimics", Eunhwa Ko, and Kevin Burgess, *Org. Lett.,* **2011**, *13*, 980-983. Copyright 2011 American Chemical Society.

Criteria for design of minimalist mimics have never been specifically delineated before. Thus, we propose the flowing four structural design criteria for minimalist mimics:

- facile syntheses with most amino acids side chains
- kinetically and thermodynamically accessible conformations for induced fit (i.e. not too rigid)
- only moderate loss of entropy on docking (i.e. only a few significant degrees of freedom that influence the side-chain orientations)
- appropriate Cα - Cβ coordinates of an accessible conformation of the mimic matching those of the secondary structure.

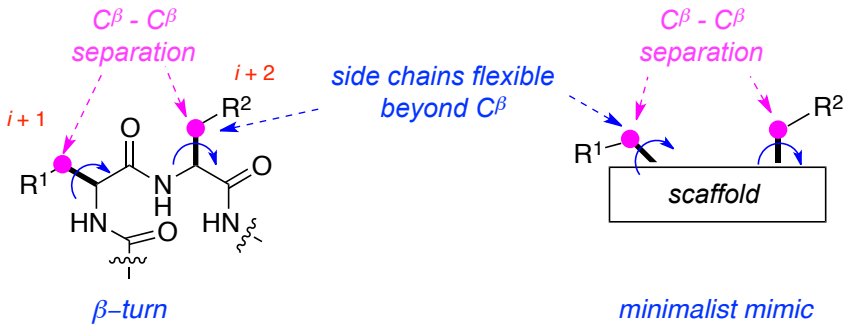
These are key parameters on the structural basis for designing minimalist mimics. Other considerations, such as water solubilities, toxicities, cell permeability and self-life are also important to design peptidomimetics. However, if the structural basis is not satisfied first, other concerns about the physical and pharmacological characteristics of the compounds might be inconsequential.

In the first criteria, Trp, Arg, Tyr, Lys and Glu are amino acids side-chains that are found frequently at hot-spots in protein-protein interactions.[7] Thus, if mimics cannot easily make with the amino acids, but they can be prepared with only simple alkyl chains, they are not good mimics. The second and third criteria are relating to the kinetic and thermodynamic accessibilities and the significant degrees of freedom.

The fourth parameter is about Cα - Cβ vector. Previously, our group proposed that Cβ - Cβ separations are critical in the design of minimalist β-turn mimics.[47,48,81] This is because the Cβ positions represent the last atoms along the side chains that are held with some rigidity in these secondary structures. Compounds **A** and **1** in Figure 5.1b were designed and synthesized as β-turn mimics for neurotrophins on this criteria, and the verification was proved with biological data; compounds **A** bind to TrkA, and compounds **1** are partial agonists for TrkC.[47,48] Cβ - Cβ distances are useful as a “rough

cut" to gauge the fit of a proposed minimalist mimic to secondary structures. A more sophisticated standard for minimalist mimics is side-chain Cα - Cβ bond vectors, because the parameter reveals how the side-chain projects into space as well as Cβ - Cβ distances. Matching Cα - Cβ bond vectors to secondary structures gives more realistic sense of the validity of the mimic. However, disadvantage for Cα - Cβ bond vector consideration is that computational work is required; it is less convenient that the intuitive approach based on Cβ - Cβ distances alone.

a



b

Figure 5.1. (a) A β-turn and general scaffolds of minimalist mimic. (b) Minimalist mimics **A** and **1** that were prepared by our group.

Minimalist mimics do not have fixed Cβ - Cβ separations (*βs* values). For example, compounds **1** have several different minimal energy conformations that correspond to preferred orientation for the two side-chains (Figure 5.2). Energy barriers for interconversion between the most contracted conformer and the most extended conformer are likely to be relatively insignificant, so this particular type of peptidomimetics could access a range of *βs* values. The extension factor is defined as *ef* = $\beta s_e/\beta s_c$, to calibrate the capacity of a minimalist mimic to span different Cβ - Cβ distances. Peptidomimetics with small extension factors can mimic a limited range of Cβ - Cβ separations, while mimics with large extension factors correspond to side-chain separations in more secondary structures. However, large extension factors are not always ideal in designs, because they might also represent excessive flexibility in the scaffold. Therefore, *ef* values are predictors of scope rather than quality.

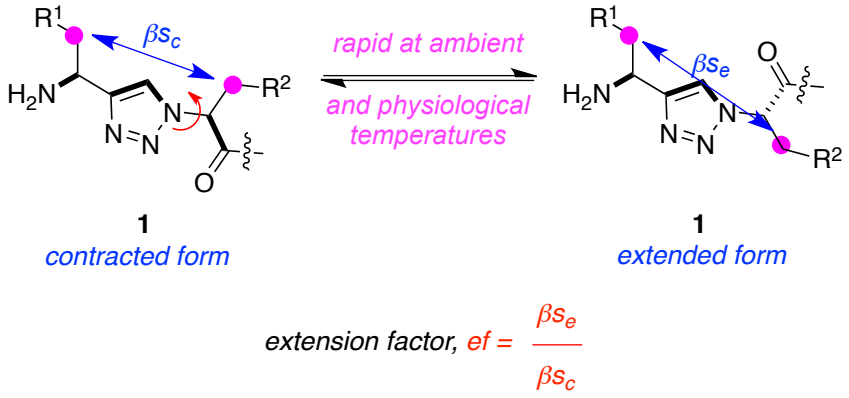


Figure 5.2. Definition of extension factor.

Six templates **1** - **6** were designed and synthesized using the structural criteria listed for minimalist mimics.[48,52,82] **1**, **2**, **5**, and **6** have piperidine or piperazine linker to facilitate assembly of these molecules into heterobivalent mimics.

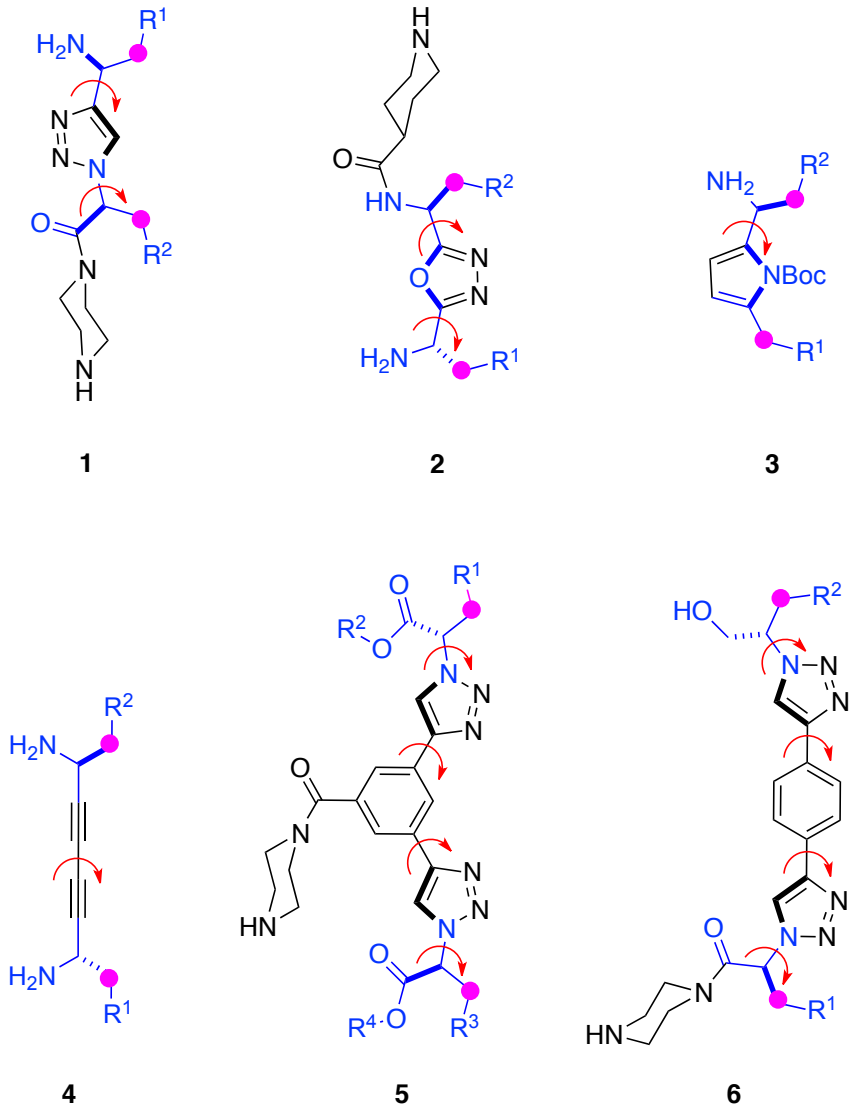


Figure 5.3. Scaffolds for universal peptidomimetics **1** - **6**.

Table 5.1 shows the extension factors deduced for mimics **A**, and **1** - **6**. Peptidomimetics **3** and **4** have the smallest *ef*, and **5** have the largest *ef* in this series. Scaffold **1** allows R^1 and R^2 side-chains to approach more closely than any other in the series of compounds. However, the *βs* value of **1** is intermediate in the series, that is, the extension factor for this mimic is not the smallest of theses compounds.

Table 5.1. Extension Factors for Peptidomimetics **A** and **1** - **6**.

compounds	βs_c (Å)	βs_e (Å)	*ef*
A	5.5	6.2	1.2
1	5.5	7.1	1.3
2	5.2	7.2	1.4
3	5.4	6.2	1.1
4	7.5	8.1	1.1
5	7.4	14.1	1.9
6	12.1	15.0	1.2

Table 5.2 shows data from modeling experiments that we performed to assess Cβ - Cβ separations for the common elements of secondary structure. It reveals considerable overlap between *βs* values for different residues in different secondary structures with the range of *βs* values accessible by each mimic. For instance, the *i* to *i*+3 *βs* in a type I β-turn is about equal to that for the *i* to *i*+2 in an inverse γ-turn (ca. 5.4 Å), so a scaffold that can access this *βs* can mimic those side-chains in both secondary structures. Indeed, the extent of Cβ - Cβ separation overlap in Table 5.2 indicates that almost all the side-chain Cβ - Cβ separations could be mimicked with relatively few peptidomimetics. In other word, small sets of such scaffolds can be designed to analogue local pairs of amino acids including noncontiguous ones in any secondary structures; they are “universal peptidomimetics”.

Table 5.2. Cβ - Cβ Distances for Mimics **A** and **1** - **6** with Secondary Structures.

structure	sequence	*βs*(Å)	**A**	**1**	**2**	**3**	**4**	**5**	**6**
ef			1.2	1.3	1.4	1.1	1.1	1.9	1.2
α-helix	*i*-*i*+1	5.2							
	i-*i*+2	7.1							
	i-*i*+3	5.6							
	i-*i*+4	6.5							
	i-*i*+5	9.8							
	i-*i*+6	10.8							
	i-*i*+7	10.7							
	i-*i*+8	12.9							
β-turn (type-I)	*i*-*i*+1	5.7							
	i-*i*+2	5.1							
	i-*i*+3	5.4							
	i+1-*i*+2	5.2							
	i+1-*i*+3	7.5							
	i+2-*i*+3	5.6							
β-sheet (parallel)	*i*-*i*+1	5.8							
	i-*i*+2	7.1							
	i-*i*+3	11.1							
	i-*i*+4	13.2							
	i-*i*'	5.5							
	i-*i*'+1	7.2							
	i-*i*'+2	9.0							
	i-*i*'+3	11.5							
	i-*i*'+4	14.4							
β-sheet (anti-parallel)	*i*-*i*+1	5.8							
	i-*i*+2	6.5							
	i-*i*+3	11.0							
	i-*i*+4	12.8							
	i-*i*'	4.5							
	i-*i*'+1	7.6							
	i-*i*'+2	8.9							
	i-*i*'+3	12.7							
	i-*i*'+4	14.8							
γ-turn (classic)	*i*-*i*+1	4.7							
	i-*i*+2	7.2							
	i+1-*i*+2	5.1							
γ-turn (inverse)	*i*-*i*+1	5.7							
	i-*i*+2	5.4							
	i+1-*i*+2	6.2							

5.2 Analyses of Universal Peptidomimetics 2, 5, and 6

For peptidomimetics **2** - **6**, density functional theory (DFT) calculations and quantum molecular dynamics (QMD) were performed. DFT calculation was used to investigate kinetic accessibility for compounds. In these calculation, Gaussian 03 was used at the B3LYP level of theory with a 6-31+G(d') basis set and a polarized continuum solvation model with a dielectric of 80. [83,84]

The DFT method facilitates calculation of transition-state energies and relative energies of resting conformations, but it does not show the all conformations that can be formed and rank their relative energies, thermodynamic accessibility. To do this, QMD was used. In this technique, a molecule is minimized and then subjected to a molecular dynamics run at high temperature (1000 K) for a short time (600 ps); 600 conformational states are recorded during this run (i.e. every 1 ps) and minimized via molecular mechanics.[85-88] The lowest energy structures below a user-defined cutoff are selected and then clustered into families on the basis of root-mean-square (rms) deviations from user-defined atoms.

Peptidomimetics **2** were synthesized by Dr Jing Liu, peptidomimetics **5** by Dr Genliang Lu, and peptidomimetics **6** by Dr Amber Schaefer and Shuhei Shimizu.[52]

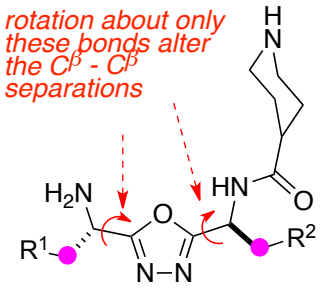
5.2.1 Analyses of 1,3,4-Oxadiazole-based Peptidomimetics 2

Pharmaceuticals including oxadiazoles scaffold have been shown to have a range of bioactivites and applications, including bactericidal, fungicidal, analgesic, anti-inflammatory, antiproteolytic, anticonvulsant, nervous system depressant, sedative and local anesthetic.[89-92] As triazole-based peptidomimetics, **A** and **1**, 1,3,4-oxadiazole-based mimics **2** have a compact five-membered heterocycle that might be ideal for the formation of universal peptidomimetics.

Compounds **2** have two significant degrees of freedom that separate the key Cβ–atoms (Figure 5.4a). Cβ - Cβ separations of **2** are between 5.2 Å and 7.2 Å (Figure

5.4b). In the Table 5.2, the compounds can become mimics of most secondary structures with the Cβ - Cβ separations.

a



2

b

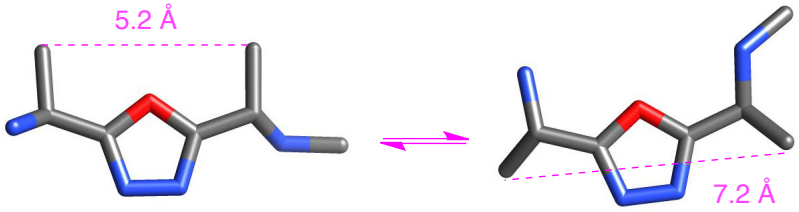
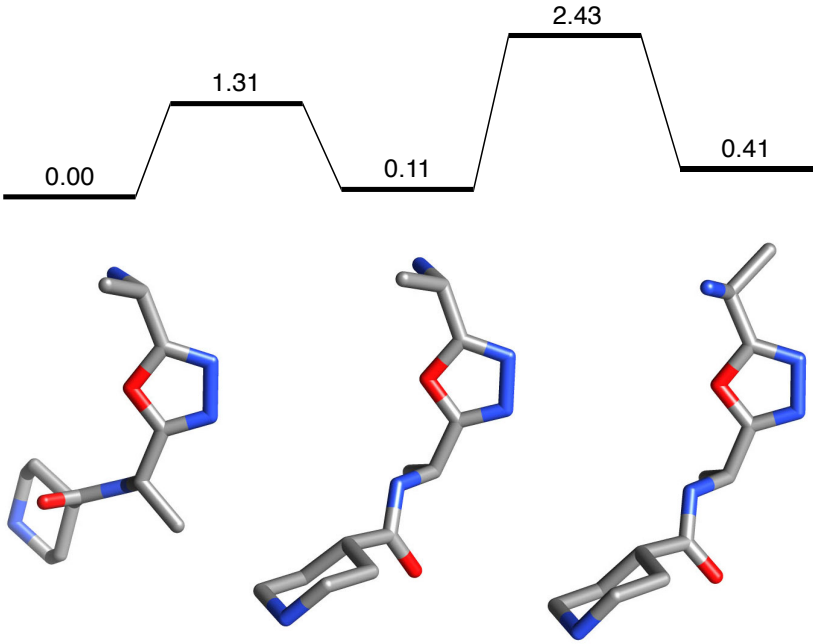


Figure 5.4. (a) Degrees of freedom for mimic **2**. (b) Conformations corresponding to the βs_c and βs_e of mimic **2** (piperidine ring omitted).

In DFT calculation for **2**, the global minimum energy structure that was identified was set to 0 kcal/mol, and is shown on the left side of Figure 5.5a. Conformer on the right side of Figure 5.5a is gotten by rotation of the bonds. The conformer is only 0.41 kcal/mol above the global minima, and the highest energy barrier to arrive at the conformer from the global minima is 2.43 kcal/mol. The energy barrier can be easily surmounted at room temperature. The conformer overlays with the *i*-*i*+3 residues in an

α-helix as Figure 5.5b. It corresponds to the results from Cβ - Cβ separations in Table 5.2; the scaffold **2** can have Cβ - Cβ separations corresponding to the *i*-*i*+3 residues in an α-helix.

a



b

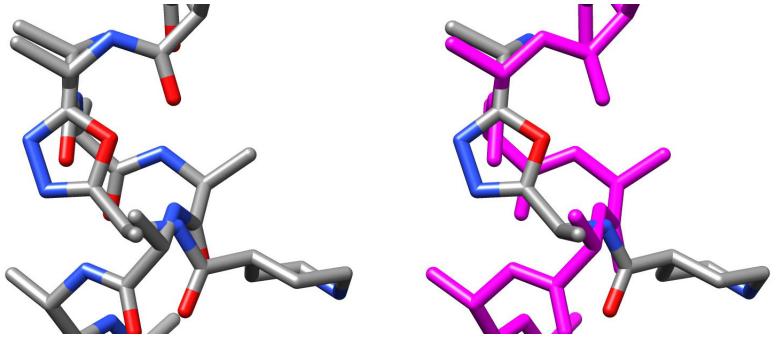


Figure 5.5. (a) Transposition of the global minima of **2** into a conformation that mimics the *i* and *i*+3 residues in an α-helix by rotation around one of the significant bonds and then the other (ΔG° in kcal/mol). (b) Overlay of the latter conformation on an ideal α - helix.

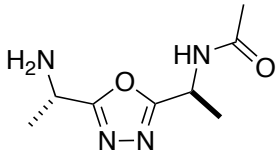
In QMD techniques for **2**, application of a 0.45 kcal/mol energy cutoff gave 180 minimized conformations. These 180 structures were clustered by overlay of the Cα and Cβ atoms within 0.3 Å RMSD. This process gave seven families of conformations (Table 5.3). Family 3 has the most structures that overlaid, and the minimum energy structure in this family was only 0.15 kcal/mol above the overall minimum energy conformation. Conformations within this family overlay well with the *i*-*i*+3 side-chains of an ideal α-helix conformation. This is the same conclusion that was reached using the DFT method (Figure 5.5b). Other structure in this family also overlay the *i*-*i*+2 side-chains of an ideal inverse γ-turn (Figure 5.6a). Moreover, structures in family 5 overlay with an ideal type I β-turn at the *i*+1-*i*+2 side-chains (Figure 5.6b).

Table 5.3. QMD Analysis of **2**.

	F1	F2	F3	F4	F5	F6	F7
population	66	21	72	11	7	1	2
ΔE (kcal/mol)[a]	0.00	0.12	0.15	0.26	0.28	0.34	0.40
minimum C^{β}-C^{β}	6.32	6.07	5.91	5.93	5.65	6.50	6.92
maximum C^{β}-C^{β}	6.93	6.32	6.48	6.53	5.92	6.50	6.92
corresponds to[b]			α-helix γ-turn		β-turn		

[a] Energy differences from lowest energy conformer. [b]Representative structures in the families highlighted overlays with the secondary structures as represented in the figures accompanying this table.

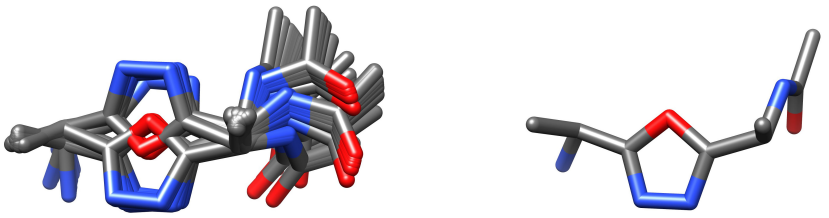
Table 5.3. continued.



2
abbreviated structure

The QMD method illustrates how conformations of peptidomimetics **1** can mimic three different secondary structures. QMD facilitates sampling of conformation space, and the DFT approaches can be used to reveal if the desired conformations are kinetically accessible.

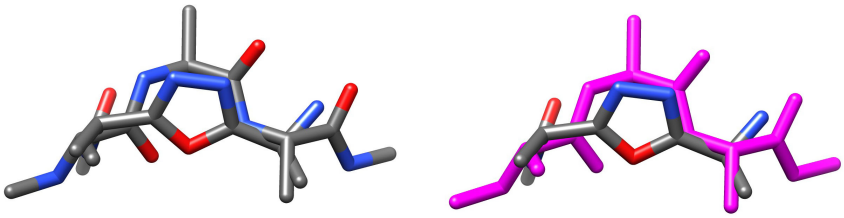
a



clustered conformations

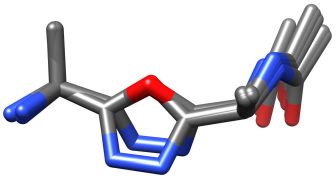
conformer with minimum energy in family

Figure 5.6. QMD data for compound **2**. (a) Data from family 3 illustrating overlay with an inverse γ-turn (overlay with an α-helix is shown in Figure 5.8). (b) Data from family 5 illustrating overlay with a type I β-turn.

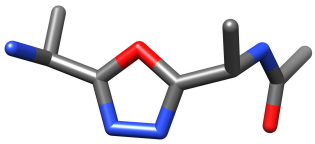


overlay of minimum energy confromaiton with an inverse γ-turn

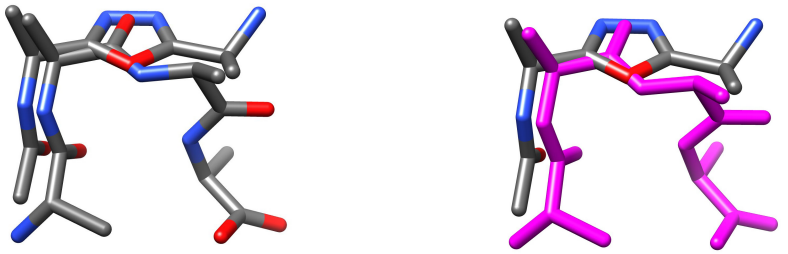
b



clustered conformations



conformer with minimum energy in family



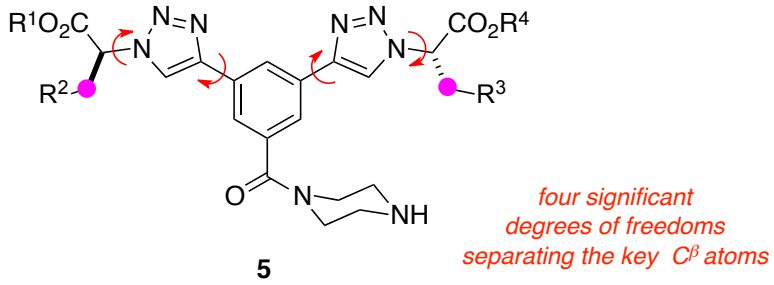
overlay of minimum energy confromaiton with a type 1 β-turn

Figure 5.6. contiuned.

5.2.2 Analyses of "Kinked" Bistriazole-based Peptidomimetics 5

Peptidomimetics **5** have four significant degrees of freedom separating the key Cβ - Cβ separations. The compounds' Cβ - Cβ separations are between 7.4 Å and 14.1 Å. The *ef*, (ca. 1.9) is the largest value in the series. From the separations, Table 5.2 shows the compounds can mimic α-helix, β-turn, β-sheet (parallel and anti-parallel), and classic γ-turn.

a



5

b

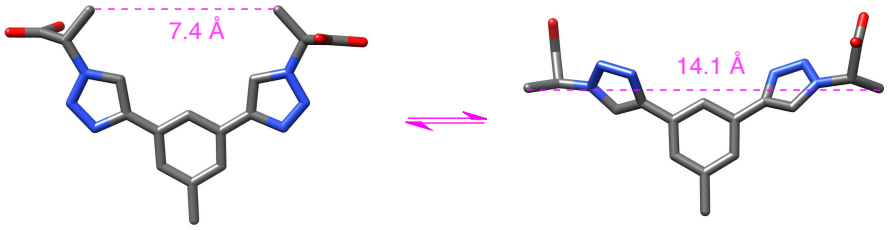


Figure 5.7. (a) Four significant degrees of freedom for compounds **5**. (b) Conformations corresponding to the βs_c and βs_e of mimic **5**.

In DFT analysis for compounds **5**, a conformer (right side in Figure 5.8b) above 0.07 kcal/mol of the global minimum conformation (left side in Figure 5.8a) overlays with "cross strand" *i*-*i*'+3 residues of anti-parallel β-sheet. The energy barrier to reach the β-sheet mimic conformer is only 4.04 kcal/mol that can be surmounted at room temperature.

a

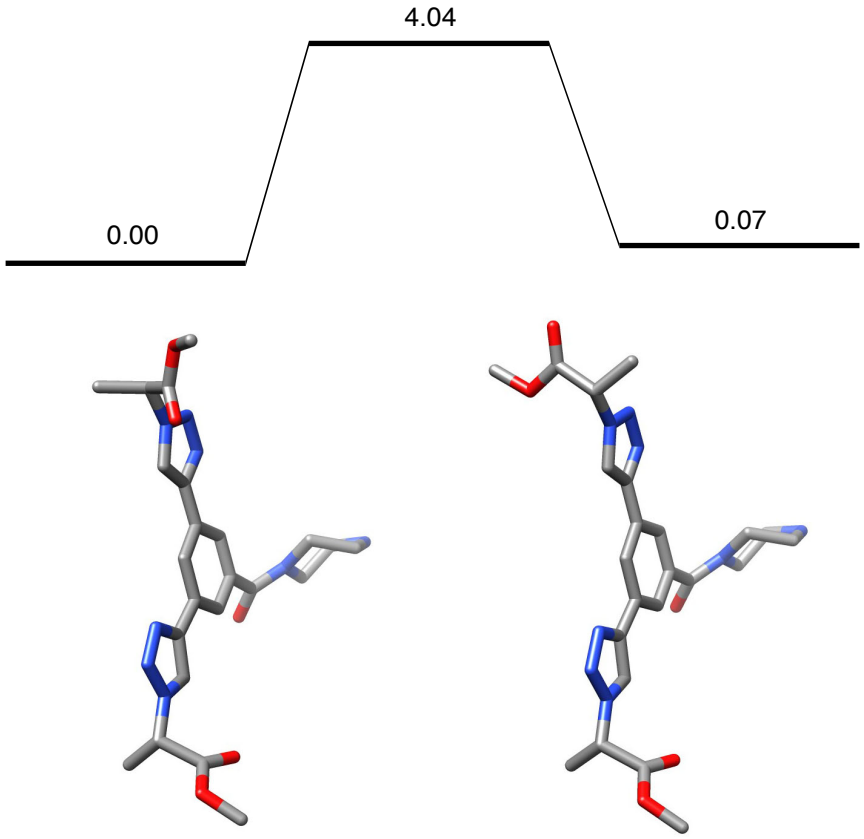
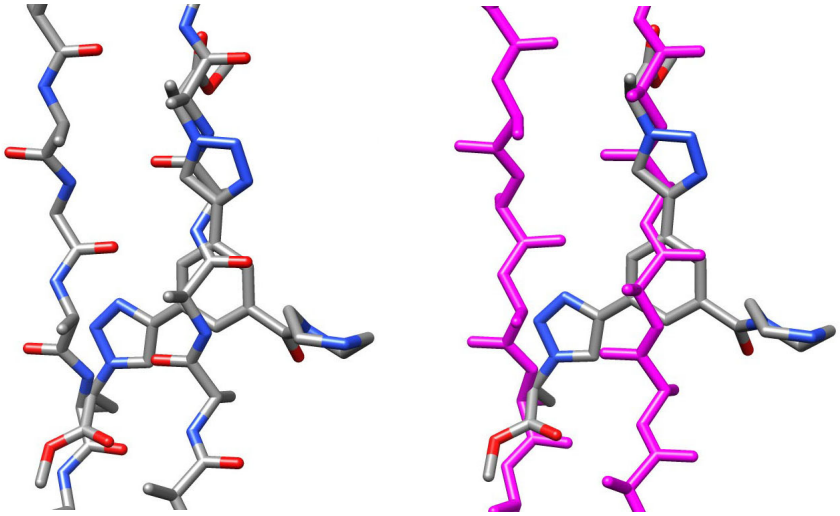


Figure 5.8. (a) Transposition of the global minima of **5** into a conformation that mimics the *i* and *i'*+3 residues in an anti-parallel β-sheet (ΔG° in kcal/mol). (b) Overlay of the latter conformation on an anti-parallel β-sheet.

b

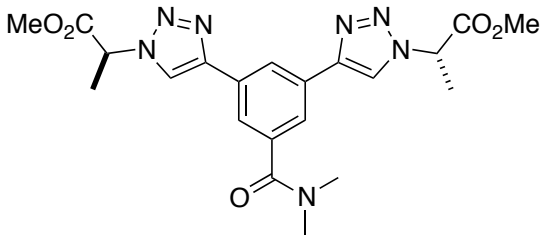


overlay of minimum energy confromaiton with anti-parallel β-sheet

Figure 5.8. continued.

The QMD analysis of compounds **5** with the abbreviated structure shows the scaffold has total 15 families within 0.3 kcal/mol energy cutoff, and conformers in family 3, 4, and 7 can become mimics for α-helix, and β-sheets (Table 5.4).

Table 5.4. QMD Analysis of **5**.[a]



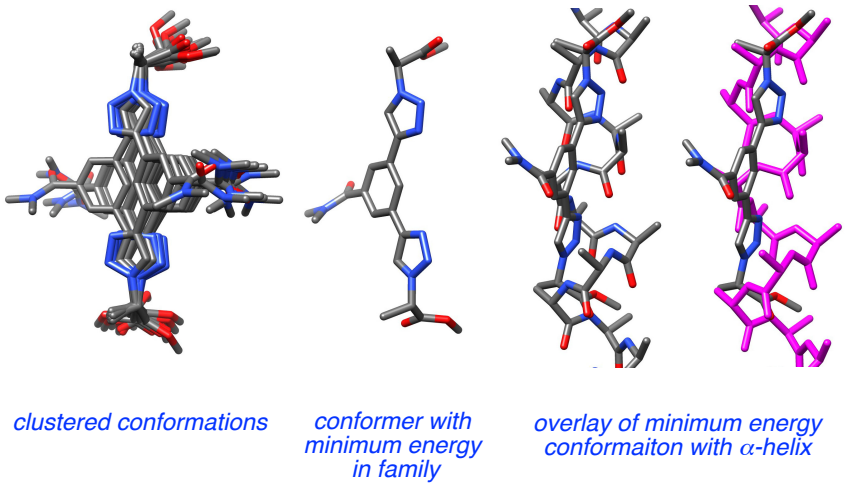
5
abbreviated structure

	F1	F2	F3	F4	F5	F6	F7	F8
population	18	17	13	8	4	11	2	7
ΔE (kcal/mol) from lowest energy conformer overall	0.00	0.01	0.03	0.07	0.08	0.09	0.13	0.15
minimum C^{β}-C^{β}	13.36	12.00	12.79	12.38	11.43	11.04	11.33	9.51
maximum C^{β}-C^{β}	13.77	12.60	13.65	12.92	11.84	11.73	11.39	10.17
corresponds to[b]			α-helix	β-sheet[c]			β-sheet[d]	

[a] Total 15 families were identified, but F9 - 15 were omitted because F9 and F12 - 15 have only one structure, and F10 and F11 have four and five structures. [b] Representative structures in the families highlighted overlays with the secondary structures as represented in the figures accompanying this table. [c] Antiparallel form. [d] Parallel form.

The conformer overlaid for anti-parallel β-sheet in DFT calculation is included in family 4. Conformer in family 3 can mimic *i*-*i*+8 residues in α-helix, and another conformer in family 7 overlays with *i*-*i*+4 residues of a parallel β -sheet.

a



b

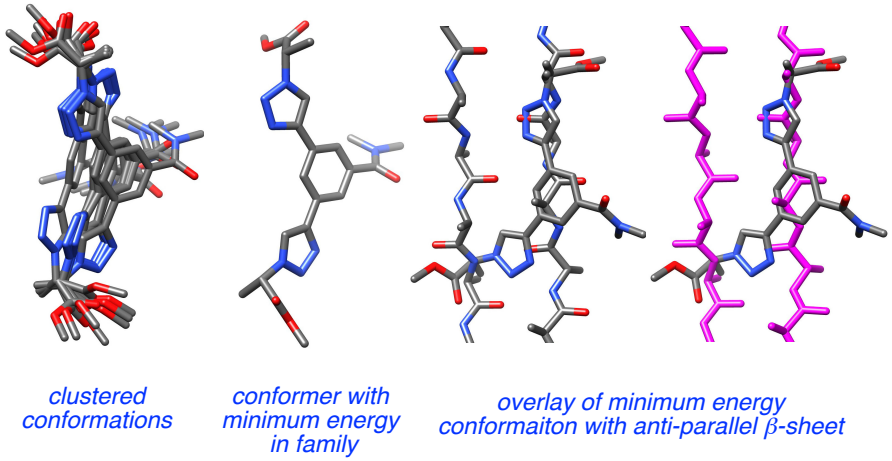


Figure 5.9. QMD data for compound **5**. (a) Family 3 overlaid with an α-helix. (b) Family 4 overlaid with an anti-parallel β-sheet. (c) Family 7 overlaid with a parallel β-sheet.

c

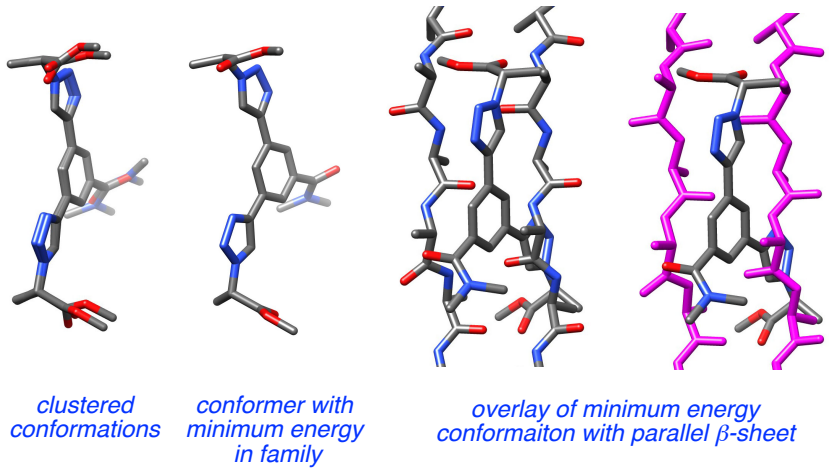
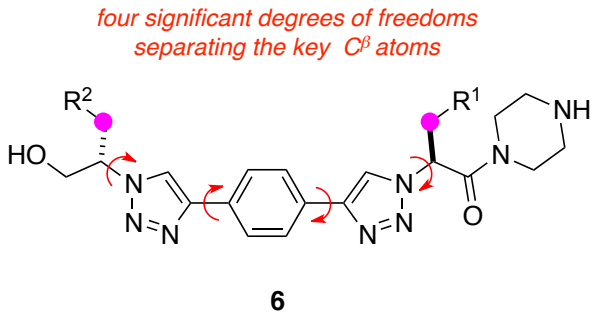


Figure 5.9. continued.

5.2.3 Analyses of "Linear" Bistriazole-based Peptidomimetics 6

Peptidomimetics **5** feature a 1,3-disubstituted benzene as the central ring. An analog 1,4-disubstituted benzene system **6** gives a greater possible span between the two pertinent side-chains up to 15 Å. However, the extension factor (*ef* 1.2) of compounds **6** is smaller than compounds **5** (*ef* 1.9). The compounds **6** can mimic distal side-chains of α-helix, parallel and anti-parallel β-sheets in Table 5.2.

a



6

b

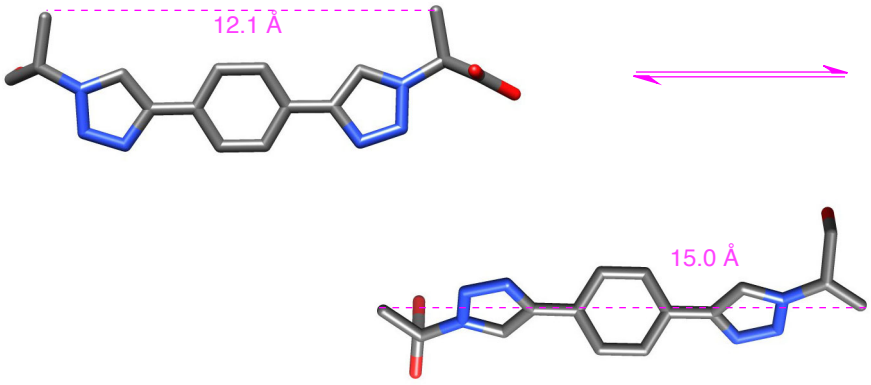
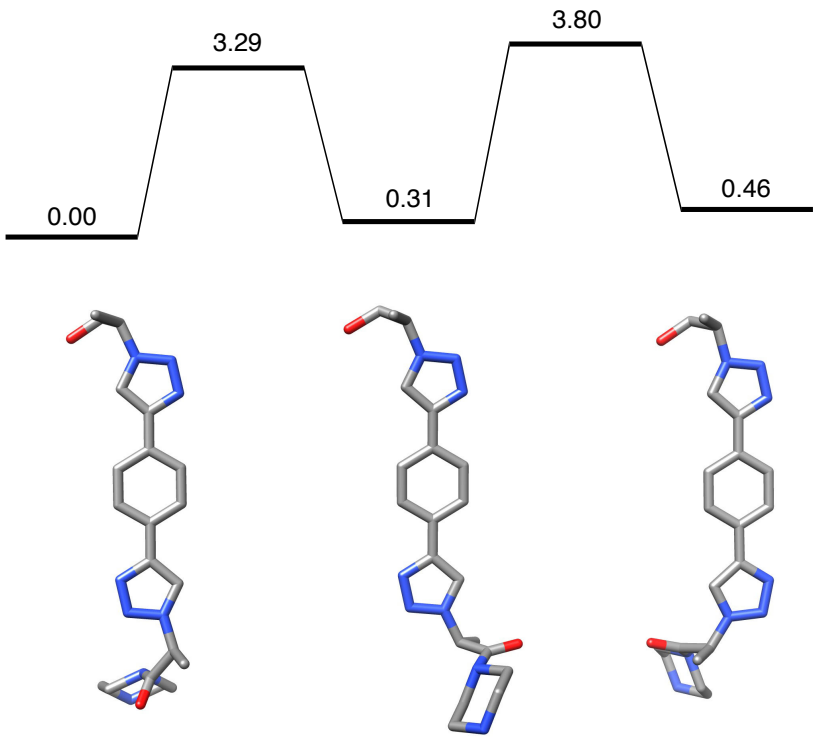


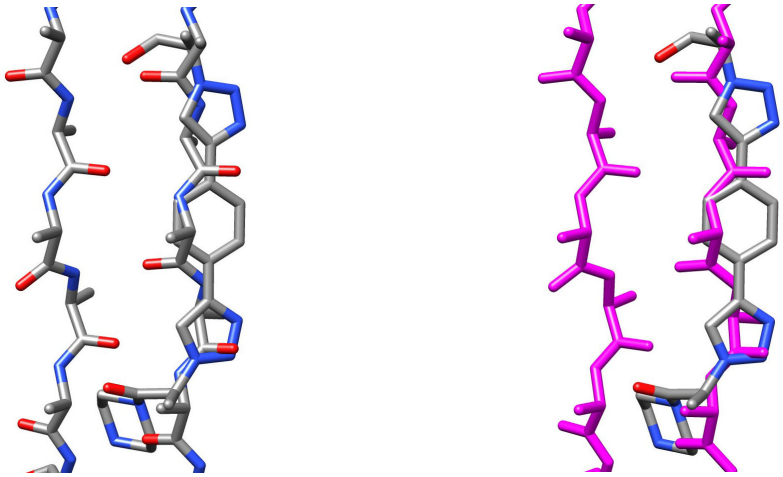
Figure 5.10. (a) Four significant degrees of freedom for compounds **6**. (b) Conformations corresponding to the βs_c and βs_e of mimic **6**.

DFT calculation shows a conformer that has 0.46 kcal/mol higher energy than a global energy minima overlays with the *i* and *i*+4 residues in a parallel β-sheet. The energy barrier to arrive at the β-sheet mimic conformer is only 3.80 kcal/mol that can be surmounted at room temperature.

a



b

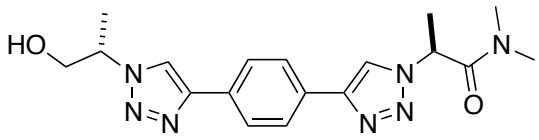


overlay of minimum energy confromaiton with parallel β-sheet

Figure 5.11. (a) Transposition of the global minima of **6** into a conformation that mimics the *i* and *i*+4 residues in a parallel β-sheet (ΔG° in kcal/mol). (b) Overlay of the latter conformation on a parallel β-sheet.

In QMD analysis, compounds **6** have 10 families within 1.0 kcal/mol energy cutoff (Table 5.5). Conformers in families 2 and 5 can be overlaid with α-helix and parallel β-sheet.

Table 5.5. QMD Analysis of **6**.[a]



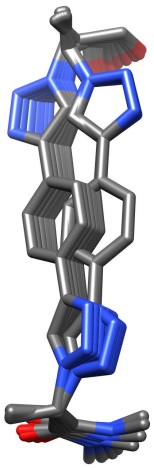
6
abbreviated structure

	F1	F2	F3	F4	F5	F6	F7
population	25	19	16	13	17	5	6
ΔE (kcal/mol) from lowest energy conformer overall	0.00	0.03	0.16	0.24	0.33	0.50	0.85
minimum C^{β}-C^{β}	13.92	13.18	14.10	13.63	12.62	13.35	13.78
maximum C^{β}-C^{β}	14.48	13.56	14.32	14.24	13.13	13.44	14.11
corresponds to[b]		β-sheet[c]			α-helix		

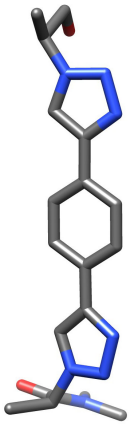
[a] Total 10 families were identified, but F8 - 10 were omitted because F8 and F10 have only one structure, and F9 has three structures. [b] Representative structures in the families highlighted overlays with the secondary structures as represented in the figures accompanying this table. [c] Parallel form.

Conformer in family 2 overlaid well with a parallel β-sheet, placing the *i*-*i*+4 at appropriate distances (Figure 5.11 and 5.12) like the result of DFT. In addition, conformer in family 5 can also mimic the *i*-*i*+8 residues of α-helix.

a



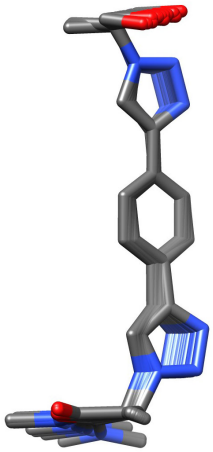
clustered conformations



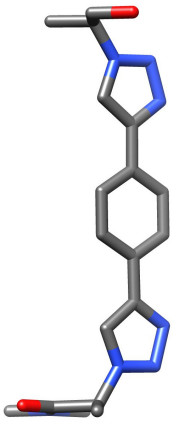
conformer with minimum energy in family

Figure 5.12. QMD data for compound **6**. (a) Family 2 overlaid with a parallel β-sheet as in figure 5.11, (b) Family 5 overlaid with an α-helix.

b

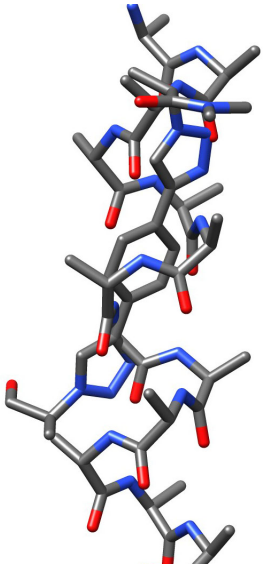


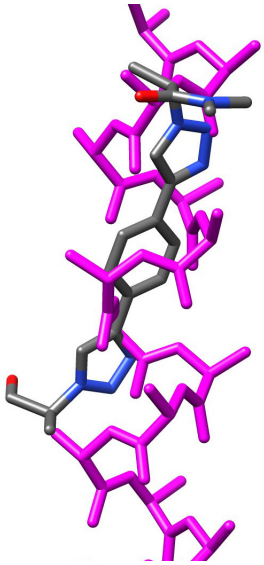
clustered conformations



conformer with minimum energy in family

c





overlay of minimum energy confromaiton with α-helix

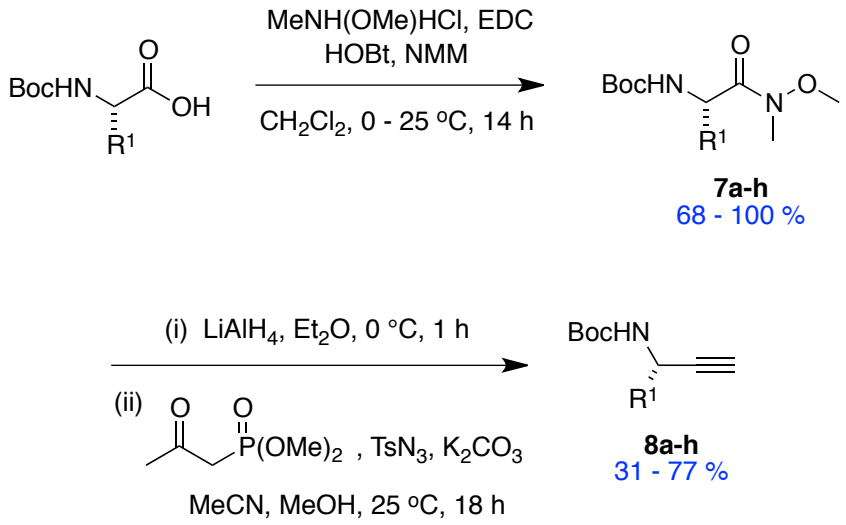
Figure 5.12. continued.

5.3 Syntheses and Analyses of 1,3-Butadiyne-based Peptidomimetics 4

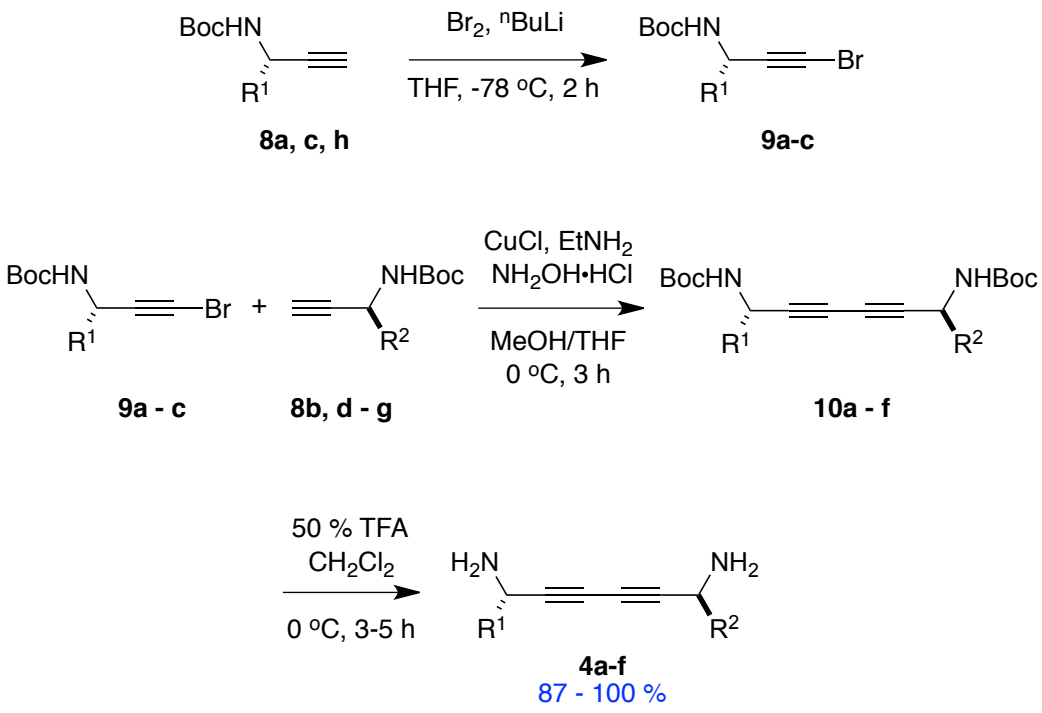
Over 1000 compounds with two or more conjugated acetylene have been isolated from Nature. These compounds show interesting biological activities in a wide area, such as antibacterial, antimicrobial, antifungal, antitumor, anticancer, anti-HIV, and pesticidal properties.[93] For example, "panazytriol" containing diyne is an important contributor to the activity of Red Ginseng. It shows inhibitory activity against MK-1 expressing tumor cells, and the cytotoxicity of mitomycin C against human gastric adenocarcinoma cell lines.[94]

1,3-Butadiyne-based compound is a suitable scaffold for minimalist peptidomimetics, because the template can be easily obtained from amino acids. Conversion of *N*-Boc amino acids into the corresponding alkyes **8a - h** was achieved via the Ohira-Bestmann modification.[95,96] The optimized condition to get the heterodiynes **10a - f** was a coupling between 1-bormo alkynes **9a - c** and the parent alkynes **8b**, **d - f**, **h** via Cadiot-Chodkiewicz condition.[94,97]

Scheme 5.1. Preparation of 1,3-Butadiyne-based Peptidomimetics **4a - f**.

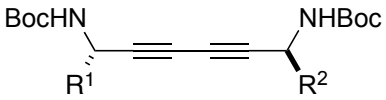


Scheme 5.1. continued.

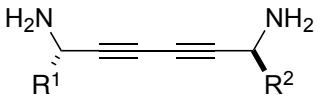


1,3-Butadiyne-base pepditomimetics **10a - f** and **4a - f** with various side-chains including Leu, Phe, Trp, Ile, Lys, Tyr, Arg, and Ser were synthesized and summarized in Table 5.6. The final compounds **4a - f** were obtained as TFA salt forms after Boc-deprotection.

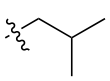
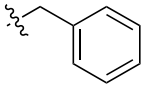
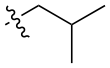
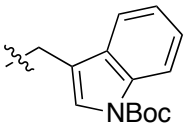
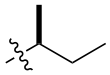
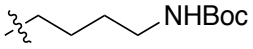
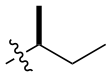
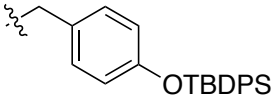
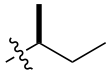
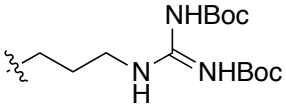
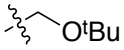
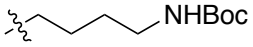
Table 5.6. Summary of 1,3-Butadiyne-based Peptidomimetics **10a - f** and **4a - f**.



10



4

compounds	R^1	R^2	**10** (%)	**4** (%)[a]
a			94	90
b			84	87
c			71	100
d			74	93
e			95	93
f			73	90

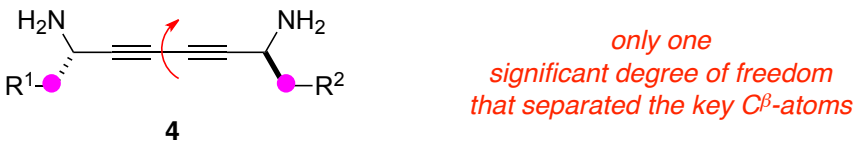
[a] TFA salt forms of **4**.

There is only one significant degree of freedom involved in moving the two Cβ atoms close to or farther away from each other and simultaneously adjusting the orientation of the Cα - Cβ vectors (Figure 5.13a). These compounds **4** are very constrained; the Cβ - Cβ separations are between 7.5 Å and 8.1 Å, and the extension factor is *ef*=1.1 (Figure 5.13b, and Table 5.1). Nevertheless, the peptidomimetics **4** can become mimics for β-turn, β-sheets, and γ-turn in limited residue sets (Table 5.2).

Figure 5.13c shows the overlay of the compound with the *i*-*i*+2 side-chains of a classical γ-turn.

The Cα–alkyne and alkyne-alkyne bonds in compounds **4** are essentially free to rotate. DFT calculations revealed that the energy maximum between the two conformations in Figure 5.13b was less than 0.01 kcal/mol, that is, the conformations were essentially equal in energy. Similarly, QMD calculations indicate that every possible conformation has almost the same energy and is equally populated. These data show that this peptidomimetics would surrender very little entropy on induced fit that sets its Cβ atoms between these ranges.

a



b

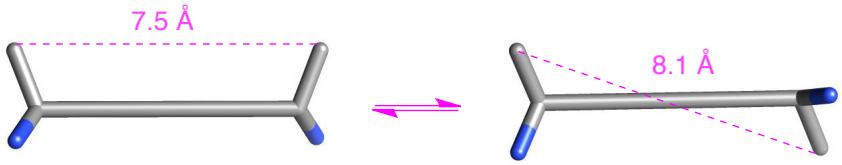


Figure 5.13. (a) Compound **4** has only one significant degree of freedom. (b) Conformations corresponding to the βs_c and βs_e. (c) Overlay of one conformation of **4** with the *i* and *i*+2 side-chains of a classical γ-turn.

c

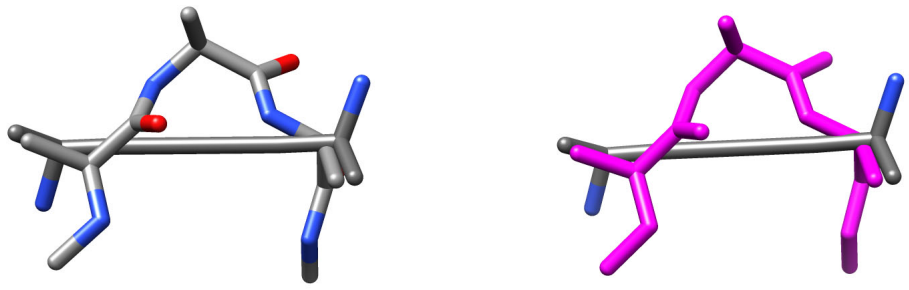


Figure 5.13. continued.

5.4 Syntheses and Analyses of Pyrrole-based Peptidomimetics 3 and 13

Pyrrole derivatives are important heterocyclic templates because they can be found in various natural products, and most compounds have biological activities. For examples, the pyrrole units are included in several natural pigments, such as heme, chlorophyll, bile pigments, and enzymes like cytochromes.[98] Thus, many synthetic approaches have been developed for construction of pyrrole building block, such as Hantsch procedure, Knorr reaction, Paal-Knorr reaction, and various metal-mediated cycloadditions.[99-101]

As amino acid surrogates, placement of a carbonyl group at the 2- or 5- position of a pyrrole arranges the *C*O and heterocyclic *N*- in a 1,3-disposition, just like the *C*O and *N* of amino acids.[102-104] There are some groups to use amino acids as starting materials for syntheses of pyrroles with peptide-like side chains.[105-108] However, none of these strategies have used two amino acids to give pyrrole derivatives with two amino acid derived side-chains.

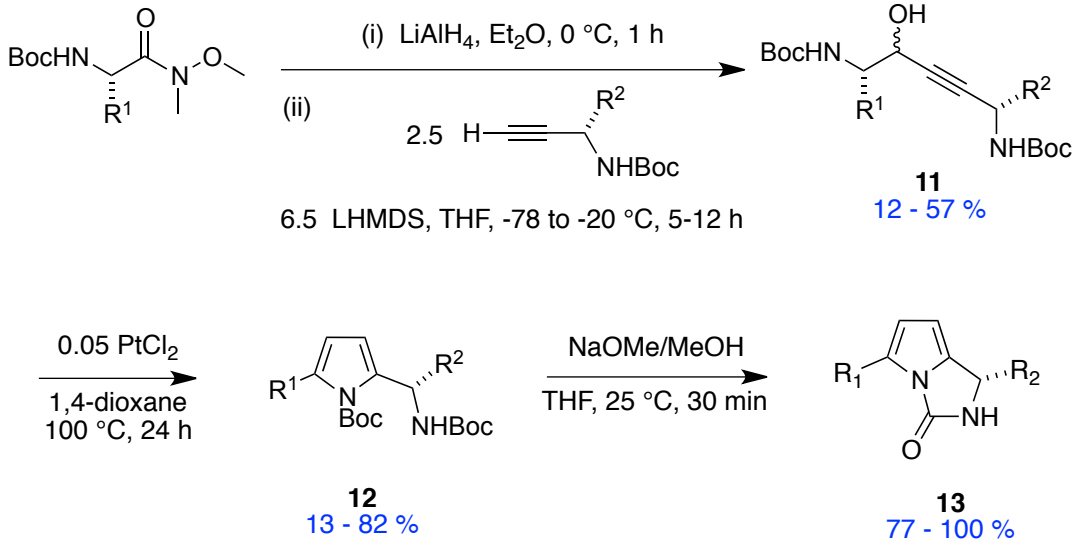
1*H*-Pyrrolo[1,2-c]imidazole derivatives that can be synthesized from pyrrole scaffolds are also interesting ring system. The templates have shown interesting biological properties as herbicidal, anti-inflammatory and anti-pyretic or as a part of

highly antimicrobial active compounds.[109-111] These pyrrole derivatives from amino acids can be universal peptidomimetics that mimic type I β-turn and inverse γ-turn.

5.4.1 Syntheses of Pyrrole-based and 1*H*-pyrroleo[1,2-*c*]imidazole-3(2*H*)-ones-based Peptidomimetics

Scheme 5.2 shows overall route to synthesize pyrrole-based and 1*H*-pyrroleo[1,2-*c*]imidazole-3(2*H*)-ones-based peptidomimetics. The syntheses of the compounds **13** began with Weinreb amides of Boc-protected amino acids.[112,113] The Weinreb amides were reduced to the corresponding aldehydes without isolation, and then immediately reacted with Boc-protected alkynes derived from amino acids in order to give propargylic (2-pyridyl)alcohols **11**.[114]

Scheme 5.2. Scheme of The Pyrrole-based Peptidomimetics **13**.

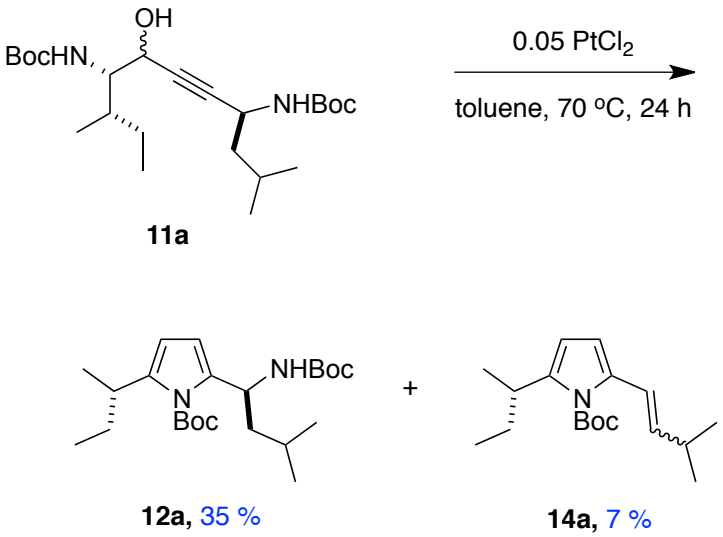


Sarpong group reported Pt(II)-mediated cyclization of propargylic (2-pyridyl)alcohols as an initial paradigm for formation of pyrrole in this work in 2007.[115]

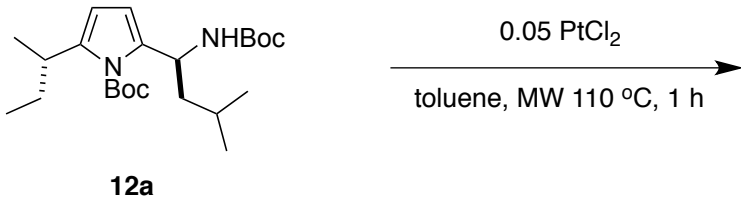
The cyclization of **11a** in the similar reaction condition with Sarpong's work gave desired compound **12a** and elimination by-product **14a** of the NHBoc group (Scheme 5.3a). To investigate an origin of the by-product **14a**, **14a** was exposed in the same cyclization condition except using a microwave. As the results in Scheme 5.3b, the by-product **14a** was drived from **12a**

Scheme 5.3. Hydroamination to build a Pyrrole Scaffold **12**.

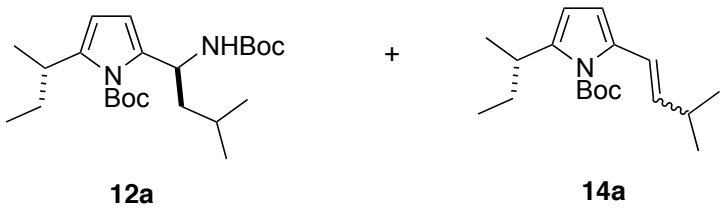
a



b



Scheme 5.3. continued.

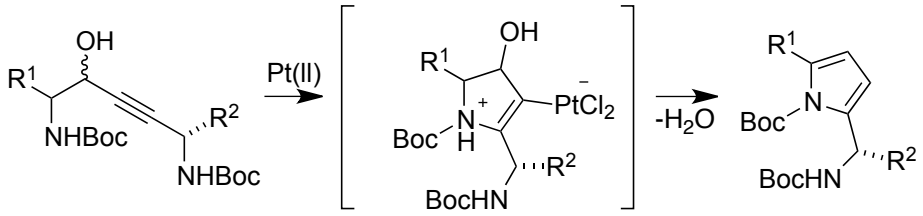


12a **14a**

4.8:1.0 by NMR

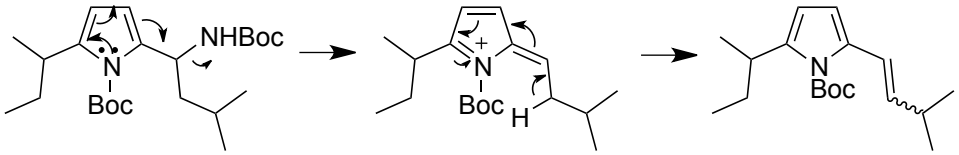
Figure 5.14a reiterates the mechanistic hypothesis outlined by Sarpong *et al*. for their transformation. The experimental data in Scheme 5.3b gave a possible mechanism for the elimination of NHBoc group as Figure 5.14b.

a



11 **12**

b



12 **14**

Figure 5.14. (a) Mechanism of Pt(II)-catalyzed hydroamination. (b) A possible mechanism of elimination of the NHBoc group.

The elimination of the NHBoc group on the newly formed pyrrole side-chains becomes a problem for pyrrole formation. Thus, a series of microwave-mediated trial reactions were run to optimize the conversion and **12**:**14** product ratio. Attempts to use other catalysts did not give better results. Similarly, acidic or basic additives, or phosphine ligands did not give benefits. However, changing the solvent from toluene to 1,4-dioxane increased the conversion, although it gave approximately the same product selectivity (Table 5.7).

Table 5.7. Pt(II)-catalyzed Pyrrole Formation through Hydroamination in Solvent Effect.

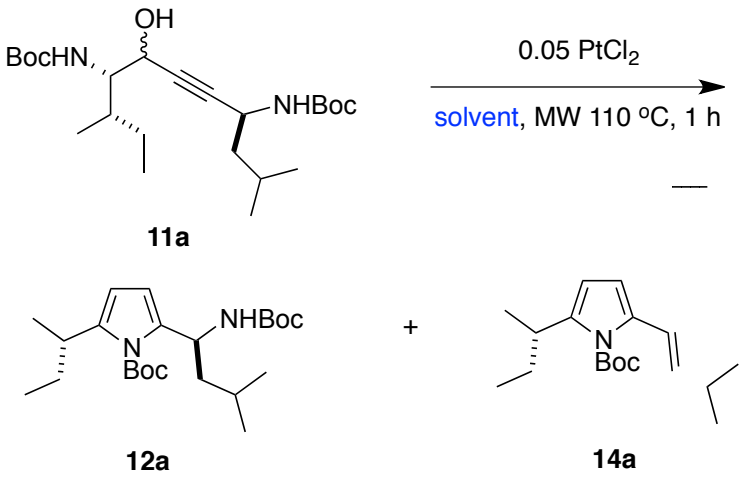
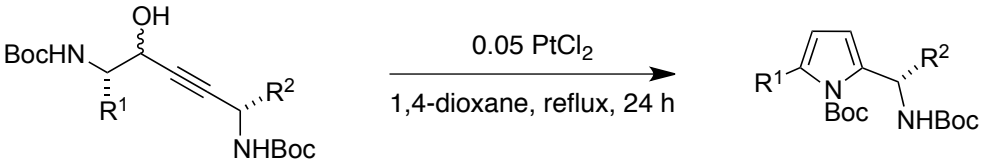


Table 5.7. continued.

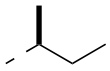
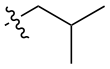
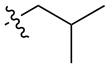
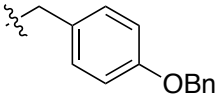
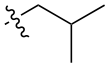
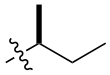
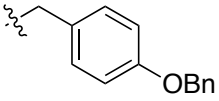
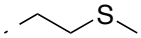
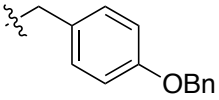
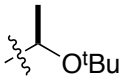
entry	solvent	**12a:14a**	conversion yield (%)
1	benzene	4.0:1.0	85
2	1,4-dioxane	4.3:1.0	80
3	THF	2.4:1.0	61
4	1,2-dichloroethane	2.1:1.0	50
5	toluene	4.8:1.0	46
6	EtOH	1.1:1.0	21

Table 5.8 shows data for a series of substrates **12** with the various amino acids side-chains, such as Ile, Leu, Val, Tyr, Met, Thr, and Gly. An adjustment had to be made for the glycine-containing substrate **11f**, because very little product was formed in this situation. Product was formed when the bis-Boc protected compound **11f'** was used, but the product yield was still low.

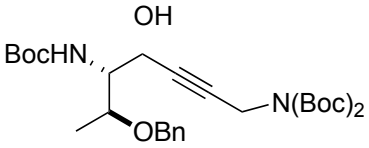
Table 5.8. Summary of Pt(II)-catalyzed Pyrrole Formation through Hydroamination.



11 **12**

compounds	R^1	R^2	**12** (%)	**14** (%)[a]
a			82	12
b			72	3
c			21	10
d			61	17
e			30	trace
f[a]		H	13	trace

[a] Substrate for **11f** was used for this reaction

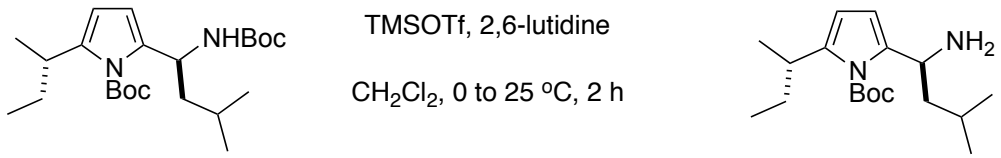


Compounds **12** have two different Boc protecting groups. Attempts to remove the two Boc groups at once failed, but the Boc protecting group on the side-chain could be selectively deprotected in TMSOTf and 2,6-lutidine (Scheme 5.4a).[116] Removal of

the pyrrole-*N* protecting group under basic conditions led to simultaneous cyclization to the 1*H*-pyrrolo[1,2-*c*]imidazole-3(2*H*)-one scaffolds **13** as in Scheme 5.4b.

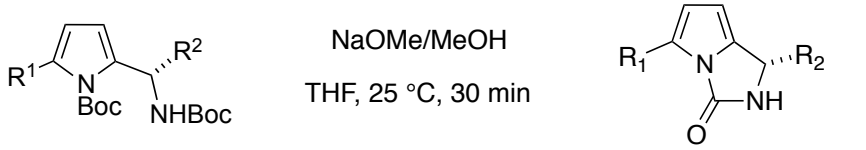
Scheme 5.4. Deprotection of Boc-protecting Groups, and Cyclization to The β-Turn Mimics.

a



12a **3a**

b

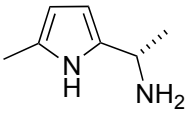


12a-e **13**

a 93 %, **b** 99 %, **c** 99%
d 77 %, **e** 94 %

5.4.2 Analyses of Pyrrole-based Peptidomimetics

Peptidomimetics **3** have only one significant degree of freedom. Cβ - Cβ separations of compounds **3** are between 5.4 Å and 6.2 Å, and the extension factor is *ef*=1.1 (Table 5.1). It is the same values to compound **4**. Although the value is the shortest value in this series, the mimic 3 can mimic α-helix, β-turn, β-sheets, and γ-turns in Table 5.2. DFT calculation shows that the maximum energy barrier as the bond rotation is 3.12 kcal/mol, that is, all conformers exist at room temperature (Figure 5.15).



3

abbreviated structure for DFT

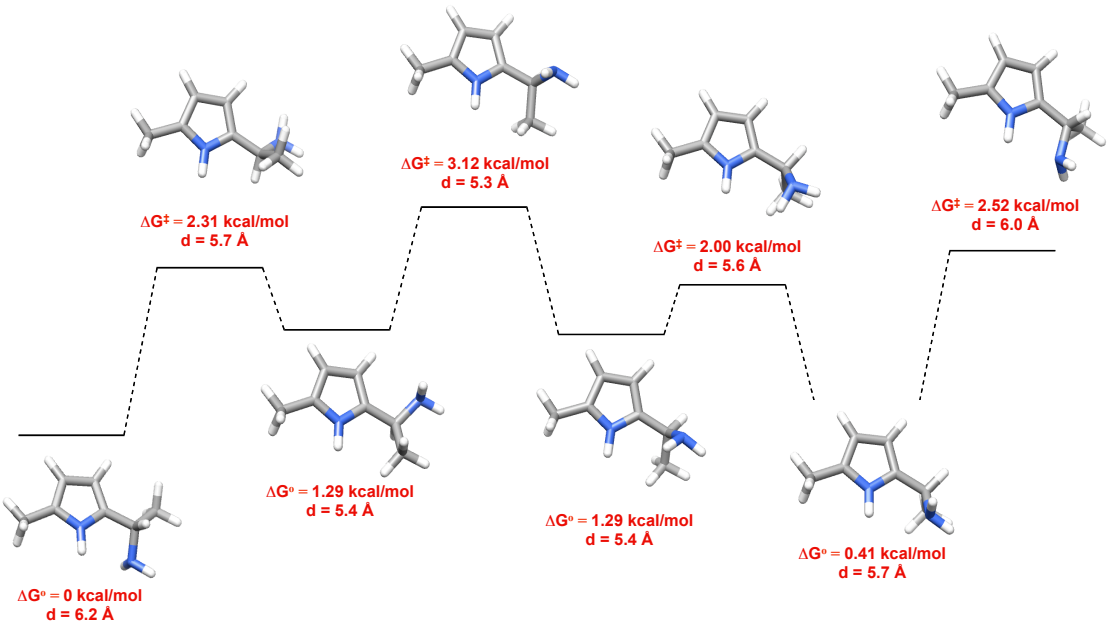
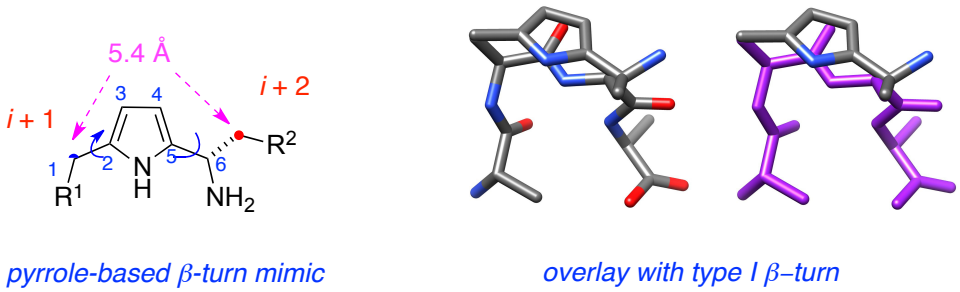


Figure 5.15. Global minimum, intermediate, and transition state structures and energy changes as the bond rotation in DFT calculation (ΔG° in kcal/mol, d=Cβ - Cβ distance).

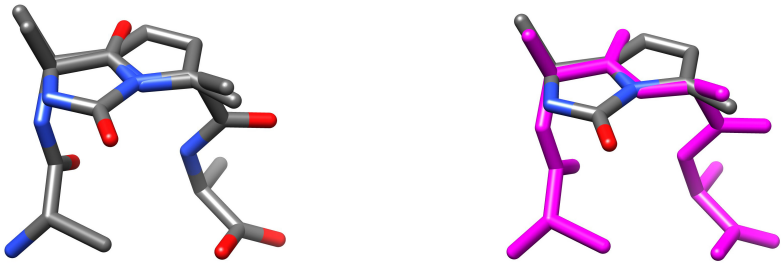
The global minimum conformer of pyrrole-based mimics overlaid with *i*+1-*i*+2 residues of type I β-turn (Figure 5.16a). However, pyrroloimidazolone-based peptidomimetics **13** show better matching with the *i*+1-*i*+2 residues of type I β-turn (Figure 5.16b). Moreover, the compounds **13** also overlaid with the *i*+1-*i*+2 residues of an inverse γ-turn.

a



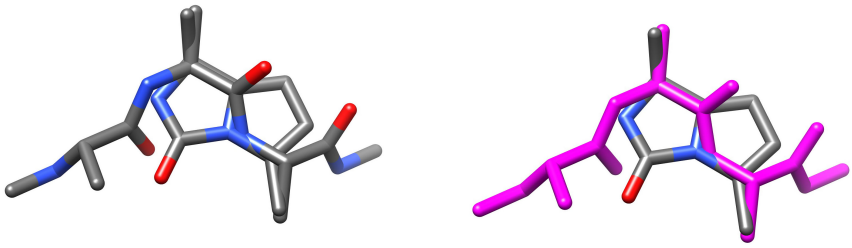
pyrrole-based β-turn mimic *overlay with type I β–turn*

b



overlay of minimum energy conformaiton with a type I β-turn

c



overlay of minimum energy conformaiton with an inverse γ-turn

Figure 5.16. (a) Overlay of pyrrol-based peptidomimetics **3** with type I β-turn. (b) Overlay of pyrroloimidazolone-based peptidomimetics **13** with type I β-turn; and, (c) with an inverse γ-turn.

5.5 Validation of Universal Peptidomimetics

Scaffolds **A** and **1** were originally designed to mimic β-turn in neurotrophins. FACS assay of compounds **A** for Trk and p75 receptors showed the compounds selectively bind to TrkA.[47] Compounds **1** are known as partial agonists for TrkC through binding, survival and neurotogenic, and signal transduction assays.[48]

Compounds **A**, **1**, **2**, and **5** were submitted to the NIH Molecular Libraries Small Molecule Repository (MLSMR) for screening. In an assay for the protein-protein interactions involving the Bcl-2 family proteins Mcl-1 and Bid, peptidomimetics **2** and **5** shows some activities. In addition, scaffold **A** mimics Bim protein in Bcl-XL/Bim interaction. In inhibition assays for PB1-domain interaction of MEK with either native or a Lys-Ala mutant of MEKK2, peptidomimetics **1** and **5** showed activities for the native MEKK2 protein.[52]

These results are very interesting, because Bcl-2 family has helical conformation in the protein-protein interactions,[117] and the interface of PB1-domain interaction of MEK has β-sheet[118]. As we expected from Table 5.1, calculations and modelings, the biology data proved that scaffold **A**, **1**, and **5** can become mimics for β-turn, α-helix or β-sheet, that is, they are universal peptidomimetics.

5.6 Conclusion

Peptidomimetics **1** - **6** were designed and synthesized as minimalist mimics. Combinations of the compounds based on minimalist mimics with complementary *βs* and *ef* values can correspond to pairs of amino acids in any secondary structure conformation. The sets of compounds call universal peptidomimetics. To prove this hypothesis, DFT calculation, QMD method, and modeling for ideal secondary structures were performed. The analysis data showed that 1,3,4-oxazoline-based peptidomimetics **2** can mimic α-helix, β-turn, and γ-turn without any significant energy barriers as the results by Cβ - Cβ separations in Table 5.1. Similarly, pyrrole-based peptidomimetics **3**, and 1,3-butadiyne peptidomimetics **4** with the smallest *ef* values in this series can mimic

close side-chains of β-turn, and γ-turn. "Kinked" and "Linear" bistrizole-based peptidomimetis **5** and **6** with large Cβ - Cβ separations are able to mimic relatively distal side-chains in α-helix and β-sheets. In addition, biophysical data also validated the hypothesis; some peptidomimetics showed significant activities for several targets with various secondary structures in targeted assays and high-throughput screening against diverse targets.

With these results, universal peptidomimetics are likely to be useful for targets where exact binding conformations are unknown, and the peptidomimetics libraries are also useful for high-throughput screening against various targets. These are because the compounds can mimic a range of secondary structures, and present any protein amino acid side-chains corresponding to hot-spots in protein-protein interactions.

In this stage, we mostly use Cβ - Cβ separations and *ef* values to design minimalist mimics because proper distances between Cβ atoms are necessary for the design. However, it is a "rough cut" to fit mimics to secondary structures. To do more accurate fitting, consideration of coordinates of the Cα - Cβ bond vectors in side-chains might be required as well as Cβ - Cβ distances. Matching Cα - Cβ bond vectors for secondary structures can be achieved with computational work. Thus, it is necessary to develop a program that can do it. In addition, so far, the designed minimalist peptidomimtics are compounds having two side-chains. However, many literatures have indicated that motifs with three side-chains play important roles in biology.[119] Increased number of side-chains or lengths may give better affinities and selectivities in protein-protein interaction. Of course, longer compounds are also important, but it would not be easy to synthesize them effectively. Therefore, in next stage, it is necessary to design universal peptidomimetics with three side-chains.

CHAPTER VI

UNIVERSAL PEPTIDOMIMETICS WITH THREE SIDE-CHAINS: DEVELOPMENT OF A Cα - Cβ VECTOR MATCHING ALGORITHM *

6.1 Introduction

Protein-protein interactions (PPIs) mediate many essential cellular processes. Misregulation of these interactions is the cause of a range of diseases,[120] hence PPIs are important for development of pharmacological probes and pharmaceuticals. Protein-protein interfaces are huge, so it is not easy to target them, [7] but it is possible, particularly by targeting "hot-spot" residues at the interface that dominate the binding free energy.[7,120-122]

Secondary structures, such as helices and strands, are the prevalent at protein-protein interfaces (Figure 6.1). For example, 40 % of secondary structures in *homo*dimeric interfaces are helical, 19 % in these are β-sheets,[123] and the remainder include turns, loops, and other. In *hetero*complexes, the percentages are a little different (helices and strands 26 % and 24 %, respectively). Consequently, secondary structures in PPIs are attractive targets since hot-spots in proteins are often associated with them.

---

*Reprinted in part with permission from "Omegatides: Constrained Analogs of Peptide Primary Sequence", Dmytro Fedoseyenko, Arjun Raghuraman, Eunhwa Ko, and Kevin Burgess, *Org. Biomol. Chem.,* **2012**, *10*, 921-924. Copyright 2011 Royal Society of Chemistry. *Reprinted in part with permission from "Pyrrolinone-Pyrrolidine Oligomers as Universal Peptidomimetics", Arjun Raghuraman, Eunhwa Ko, Lisa M. Perez, Thomas R. Ioerger, and Kevin Burgess, *J. Am. Chem. Soc.,* **2011**, *133*, 12350-12353. Copyright 2011 American Chemical Society.

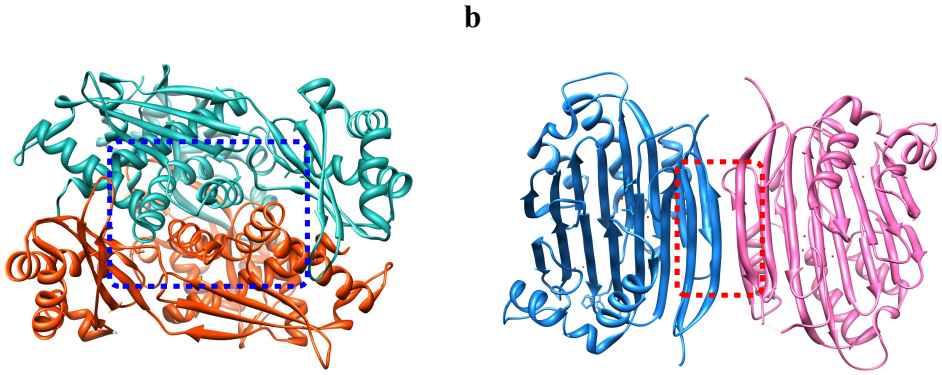


Figure 6.1. Protein-protein interface involving (a) helix-helix (PDB ID 1t02); and, (b) strand-strand (PDB ID 1jl0) interactions.

Previous chapters in this thesis have explained why our group is interested in mimicry of side-chains in secondary structures, and compounds that we call "minimalist", and "universal" mimics.[52,82] In that work, we used Cβ - Cβ separations to give a "rough cut" to fit mimics to secondary structures. However, Cα - Cβ bond vectors are a more sophisticated standard because they depict how the side-chains project. Computational methods are required to match Cα - Cβ bond vectors for mimics and ideal secondary structures. This chapter describes how we developed such a computational method and applied it to analyze the value of structures we call "omegatides" and "pyrrolinone-pyrrolidine oligomers", as new universal peptidomimetics (Figure 6.2).

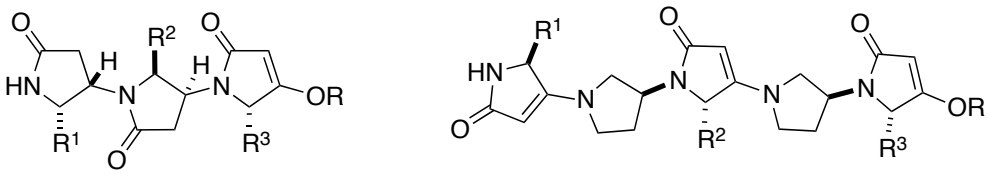


Figure 6.2. (a) Omegatides. (b) Pyrrolinone-pyrrolidine oligomers.

Omegatides and pyrrolinone-pyrrolidine oligomers have several attributes compared with our "first generation mimics" that were discussed in Chapter 5. Notably, they can express *three* side-chains and not two as before thereby giving the potential to increase affinity and selectivity of binding. Scaffolds with three side-chains roughly correspond to tripeptides; strands of three amino acids have been recognized minimal motifs for effective molecular interactions.[119] For example, Reynold's group reported that good ligand affinities are obtained from 30 heavy atoms and noted the average number of heavy atoms in three natural amino acids is 25.[124] Moreover, using an informatics method DILIMOT (Discovery of Linear MOTifs), Russell's group reported small segments of 3-10 residues play critical roles in protein interactions.[125] Thus we reasoned it was important to develop a method to compare preferred conformations of peptidomimetics with three side-chains with ideal secondary structures in terms of Cα - Cβ bond vectors. Here we illustrate this procedure and data obtained for omegatides **1** (synthesized by Dr Dmytro Fedoseyenko), and pyrrolinone-pyrrolidine oligomers **2** (prepared by Dr Arjun Raghuraman).

6.2 Development Of A Cα - Cβ Vector Matching Algorithm

Quenched molecular dynamics (QMD), described in Chapter 5, provides a method to identify and cluster preferred peptidomimetic conformations. Critically, *we elected to use the Cα and Cβ coordinates as a basis for the clustering routine*. A cut-off of 0.5 Å RMSD based on Cα and Cβ coordinates was used to identify conformers that belong in the same family (or cluster). A cluster that contained conformers within 3.0 kcal/mol of the global minimum was considered for matching with the secondary structures. This part of the procedure was developed in collaboration with Dr Lisa M. Perez as mentioned. However, an algorithm to compare these conformations with ideal secondary strucutres also had to be developed to fulfill our goal. That algorithm was developed in conjuction with Dr Thomas Ioeger of the Computer Science Department at Texas A&M University.

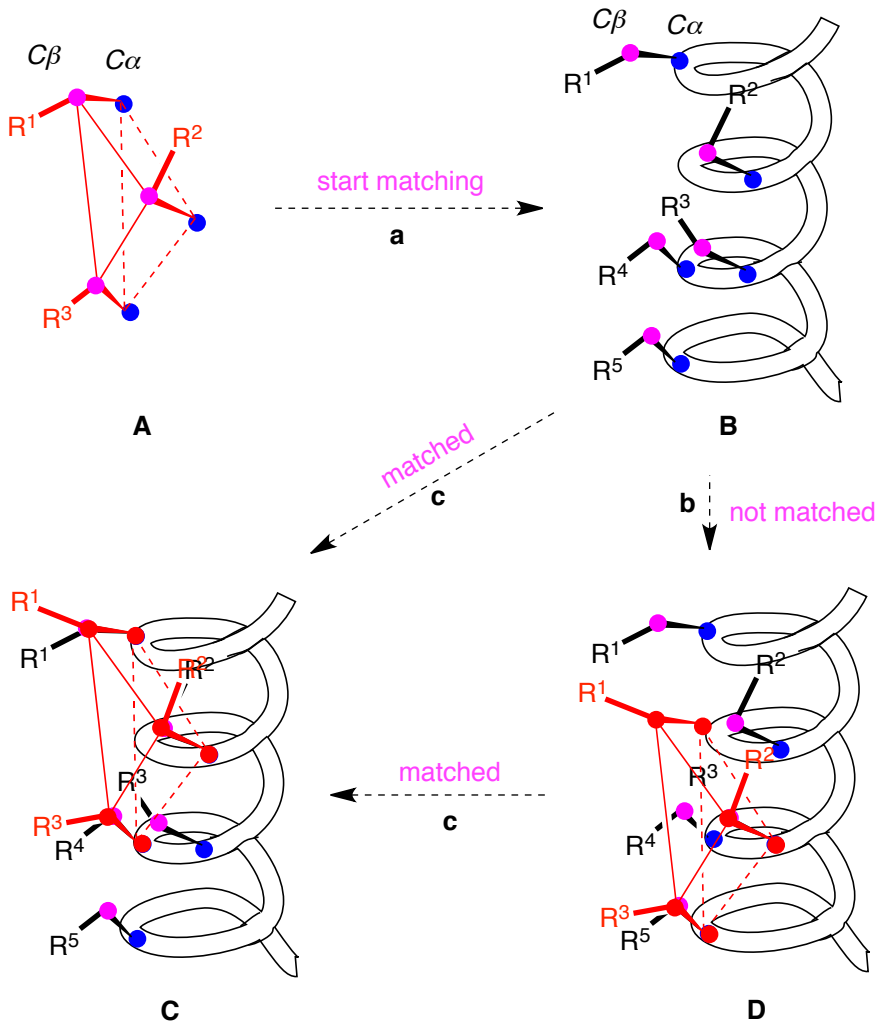


Figure 6.3. (a) Three Cα - Cβ vectors A on a preferred conformer to be evaluated for goodness of fit on side-chain residues of a secondary structure B. (b) Mismatched residue sets are disregarded. (c) Matched residue sets are evaluated in terms of RMSD and score.

The "matching algorithm" first identifies all combinations of three amino acids on inputted secondary structures. It then superposes each conformer of a mimic onto each combination of three amino acid side-chains based on Cα - Cβ vectors (Figure 6.3). Each superposition in the matching procees is scored for goodness of fit based on RMSD of Cα and Cβ atoms: this is a valuable metric because RMSD scores are widely

appreciated in science. Another metric used here is a "score" system we developed based on work of others.[126] Here the "score" is a weighted combination of distance similarities between Cα's in the two structures, and the angles between of the Cα - Cβ; like RMSD, a lower score means a better fit, but, unlike RMSD values, scores do not have units.

This program was applied to omegatides **1** and pyrrolinone-pyrrolidine oligomers **2** with *three* side chains (six Cα and Cβ coordinates) to evaluate fit of preferred secondary structures on ideal secondary structures. It can also be applied to compounds with two side-chains, in which case it uses *four* coordinates from *two* Cα - Cβ vectors.

6.3 Analyses of Omegatides as Analogs of Peptide Primary Sequence and Application of The Cα - Cβ Vector Matching Algorithm to Omegatides

Our group designed contiguous constrained amino acid surrogates that have side chains to reflect the primary sequence di- and tri-peptides; we call this design "omegatides".[127] This oligomer is based on tetramic acid building blocks wherein ϕ and ψ angles are locked, and rotation about the ω-bond is less constrained, opposite to peptides where the ω-vector is more constrained than ϕ and ψ.

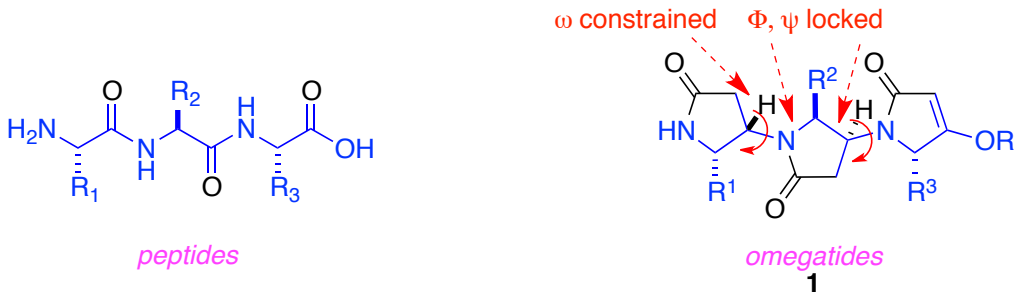


Figure 6.4. Omegatides as analogs of peptide primary sequence.

QMD and DFT calculations were performed to assess the conformational biases **1** ($R^1 = R^2 = R^3 =$ Me). QMD for **3** showed an overwhelming preference for two conformations shown in Figure 6.5a; 594 of the 600 structures existed as these two conformers to within 3.0 kcal/mol, and RMSD 0.5 Å. Members of this family present the two methyl side-chains on the same side of the molecules, so we refer to this as *syn*-**3**, and the other preferred family as *anti*-**3** because the side-chains are approximately on opposite faces. DFT calculations on **3** indicated the energy difference between the *syn* and *anti* conformers was only 0.42 kcal/mol and the energy barrier that must be surmounted to interconvert them was only 9.58 kcal/mol. That is, most conformers exist at room temperature.

a

Figure 6.5. (a) Preferred conformations for QMD. (b) Low-energy conformers and energy barrier for interconversion from DFT calculations for compound **3**. All energy shown is free energy (ΔG°) in kcal/mol.

b

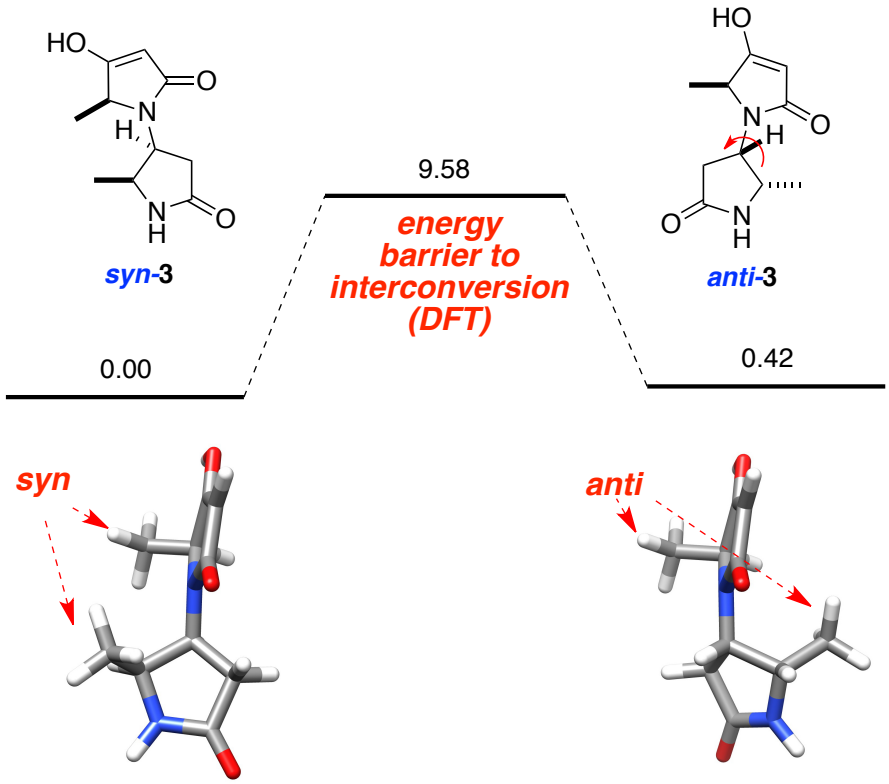


Figure 6.5. continued.

Next, we generated a Ramachandran plot for 600 conformers of different stereomers **3a - d** (LL, LD, DL, and DD; Figure 6.6). Dihedral angles of all omegatide stereomers were analyzed; most had conformations corresponding to β-sheet regions but with ϕ and ψ angle variances that are much less than in peptides. Full detail of analyses materials and methods is given in Appendix F.

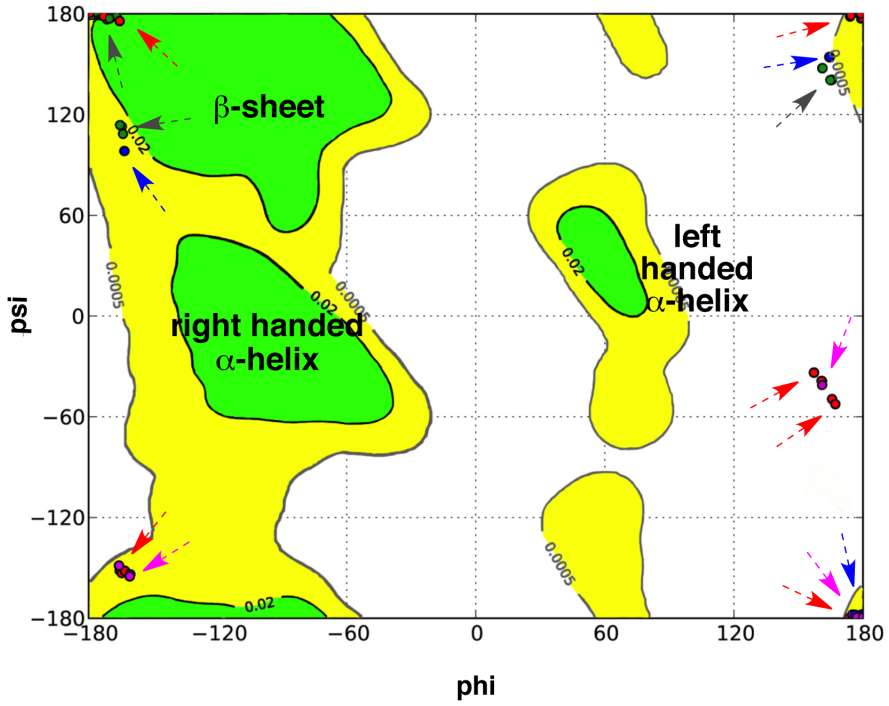


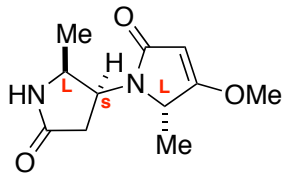
Figure 6.6. Ramachandran plot for diastereomers of **3**. Red (**3a**, LL), green highlighted with black arrow (**3b**, DL), pink (**3c**, LD), and blue (**3d**, DD).

Finally, we ran the Cα - Cβ vector matching program with **1a - c**, and **3a - b** for six different ideal secondary structures: 3_{10}-helix, α-helix, π-helix, β-strand, parallel β-sheet, and sheet-turn-sheet (*ie* an anti-parallel β-sheet with type II β-turn). In our experience, RMSD of 0.3 Å or less is a good fit for structures based on two side-chains. On the basis of this standard, data for mimics with two side-chains **3a - b** showed these compounds have better fits for strands than for helices, and the L,D-isomer **3b** is a better mimic for the strands than the L,L-isomer **3a** (Table 6.1); **3b** overlays especially well on a β-strand (RMSD 0.3 Å, score 7.8) and parallel β-sheet (RMSD 0.23 Å, score 4.3).

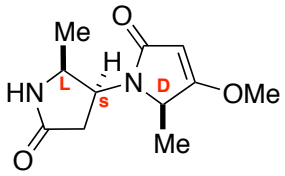
It is harder to fit six coordinates than four, so matches of the mimics involving three side-chains would have higher RMSD. To fit these structures, we used a standard of RMSD of 0.5 Å or less. As was seen in the mimics with two side-chains, the mimics with three side-chains **1a - c** showed better matching for strands than for helices.

Interestingly, the L,D,L-form **1b,** and the L,D,D-form **1c,** gave a better fit strands than the L,L,L-form **1a**.

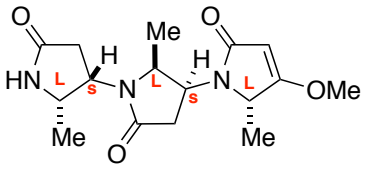
Table 6.1. Evaluation of Preferred Conformers of Mimics **3a** - **b** and **1a** - **c** on Secondary Structures.



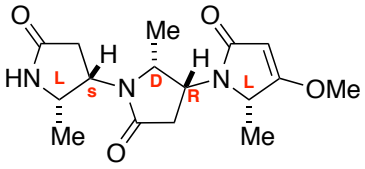
3a



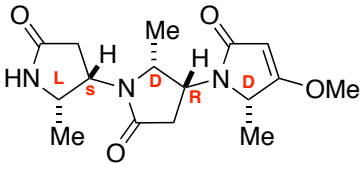
3b



1a



1b



1c

com'd	3_{10}-helix		α-helix		π-helix		β-strand		parallel β-sheet		sheet/turn-sheet	
	R[a]	S[b]	R	S	R	S	R	S	R	S	R	S
3a	0.52	21.8	0.49	22.5	0.55	26.7	0.52	12.6	0.39	14.2	0.49	14.2
3b	0.51	22.8	-[c]	-	0.66	33.6	0.3	7.8	0.23	4.3	0.31	5.0
1a	1.08	44.3	1.04	61.4	0.90	47.1	0.72	28.7	0.72	24.5	0.73	21.8
1b	1.00	29.1	0.88	42.2	0.82	54.7	0.43	17.5	0.41	15.4	0.49	16.4
1c	1.14	40.8	0.95	51.1	0.82	40.5	0.64	19.2	0.58	27.6	0.69	17.4

[a] R means RMSD (Å). [b] S means score. [c] Matching values of **3b** for α-helix were not obtained.

Figure 6.7 shows overlays of **1a** and **1b** on a parallel β-sheet. **1a** overlays on a parallel β-steet with an RMSD of 0.72 Å and a score of 24.5, and the conformer overlaid is 1.33 kcal/mol above the lowest energy structure obtained in the QMD experiment for **1a**. The L,D,L-isomer **1b** is matched for parallel β-sheet with RMSD 0.41 Å and a score 15.4. The conformer overlaid has 0.67 kcal/mol higher energy than the lowest energy structure. Overall, omegatides show better fits for strands than for helices, and this is supported by their populations in the Ramachandran plot in Figure 6.6.

a

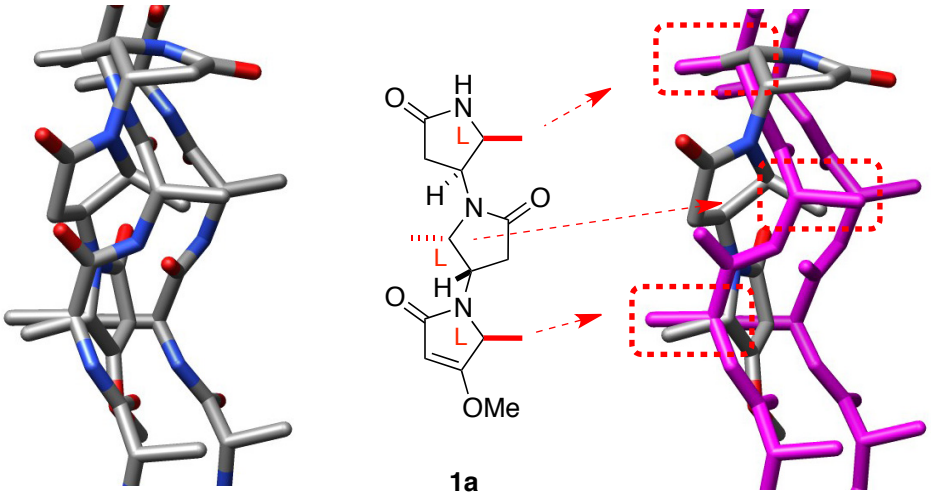


Figure 6.7. Overlays of preferred conformations of **1a** and **1b** on an ideal parallel β-sheet.

b

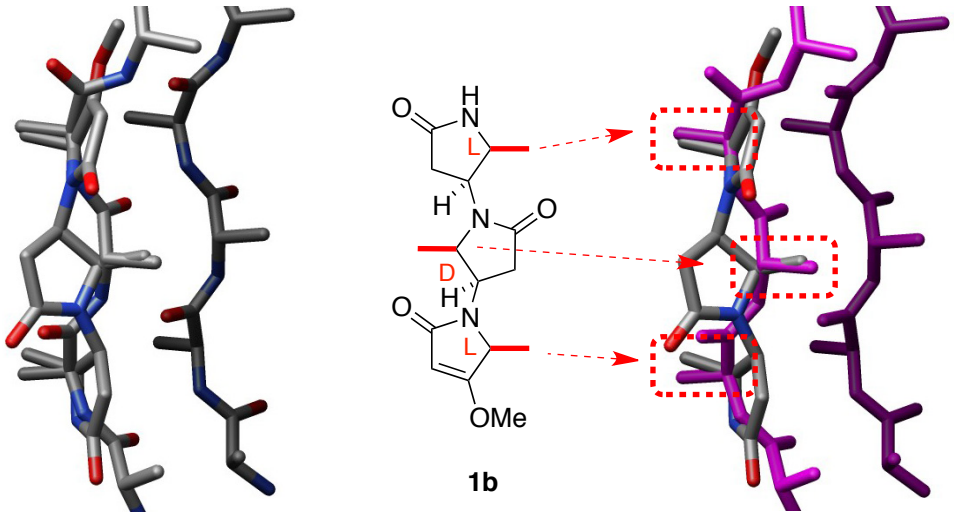


Figure 6.7. continued.

6.4 Analyses of Pyrrolinone-pyrrolidine Oligomers as Universal Peptidomimetics and Application of The Cα - Cβ Vector Matching Algorithm to Pyrrolinone-pyrrolidine Oligomers

Pyrrolinone-pyrrolidine oligomer **2** is a non-contiguous analog of omegatides **1** (Figure 6.8).[128] The templates have a variety of attributes for pharmacological probes and pharmaceutical leads. Two of their four significant degrees of freedom involve vinylogous urea bonds that are somewhat rigid due to π-*N* conjugation. None of the nitrogen atoms are basic, so the molecules are not polycationic under physiological conditions. The scaffold has some intrinsic water solubility modulated by R^1, R^3, R^5, R^C and R^N. Variation of R^C and R^N will not significantly affect the preferred template conformations.

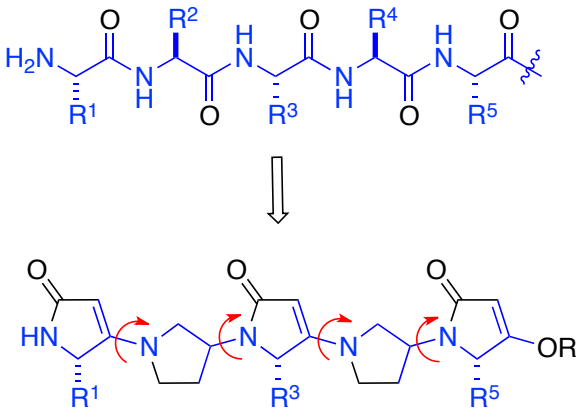


Figure 6.8. Pyrrolinone-pyrrolidine oligomer as close analogs of peptides.

We checked the kinetic accessibility of conformers via DFT calculations. The calculation gave three major intermediates, and their energies are slightly different (Figure 6.9a and b, E = the lowest energy structure, ΔG^{o} of F = 0.64 kcal/mol, ΔG^{o} of G = 1.20 kcal/mol). The maximum energy barrier is 5.10 kcal/mol; this can be surmounted at ambient temperature to interconvert the conformers. The free energy differences from the DFT calculations predict that relative populations of the three conformers (**E**, **F**, and **G**) exist in the ratio 1.00:0.34:0.12. Application of the QMD method to check thermodynamical accessibilities gave the three families, **H**, **I**, and **J**, shown in Figure 6.9c; 581 of the 600 structures existed as these three conformers to within 3.0 kcal/mol and RMSD 0.5 Å. The structure populations for each family are **H**=380, **I**=78, and **J**=123.

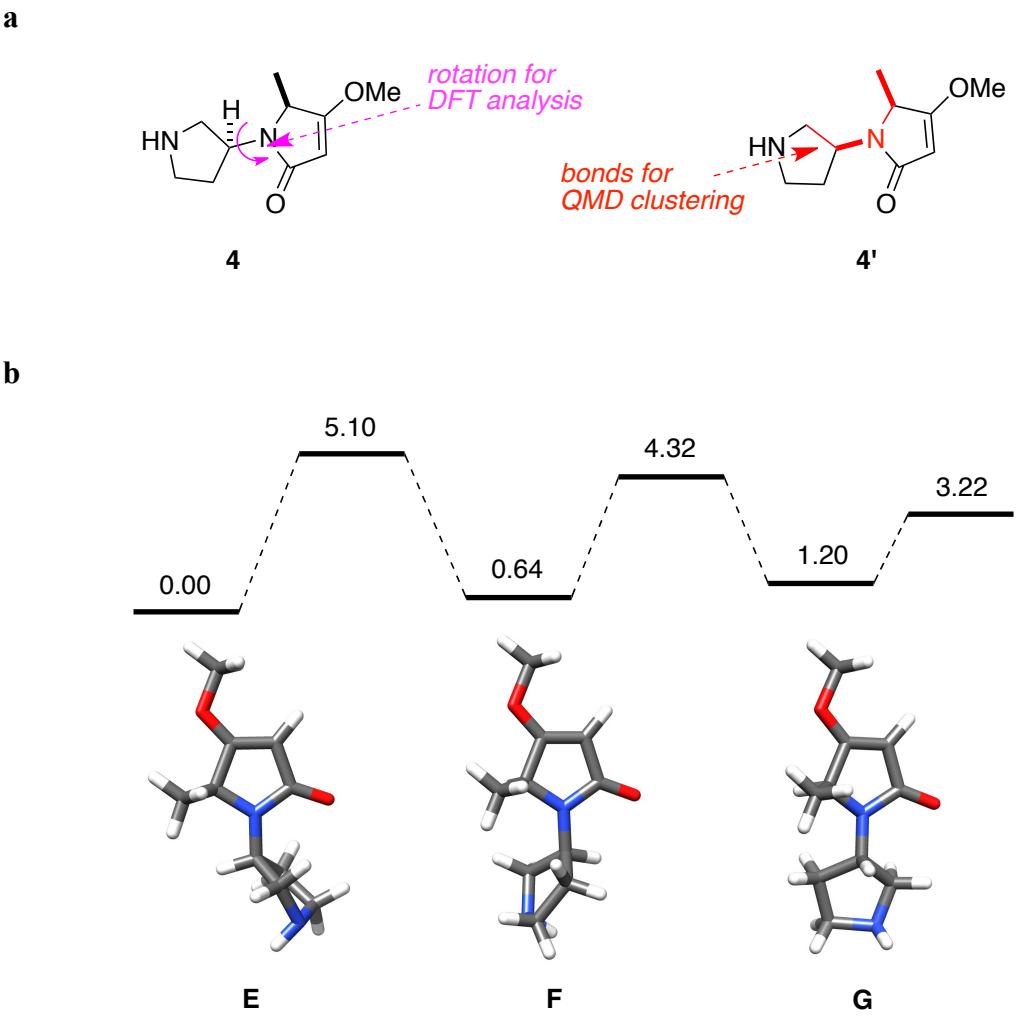


Figure 6.9. (a) Structures and parameters used for DFT and QMD analyses. (b) Low-energy conformers and energy barriers for interconversion from DFT calculations, and all energies shown are free energies (ΔG^{o}) in kcal/mol. (c) Preferred conformers from QMD calculations.

c

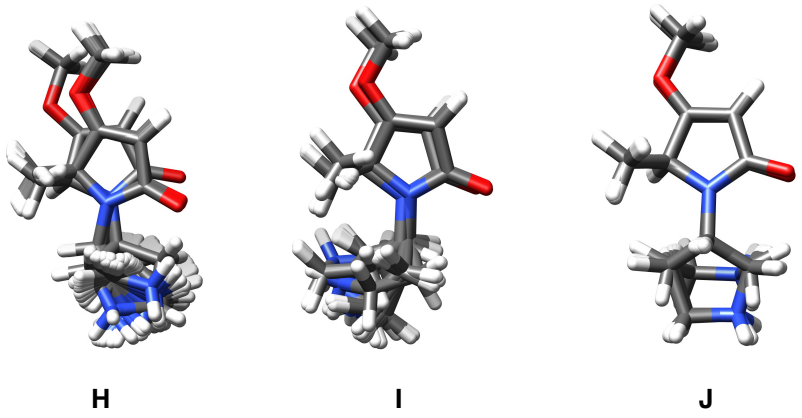
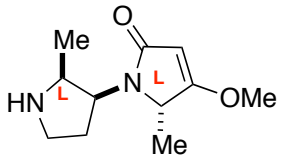


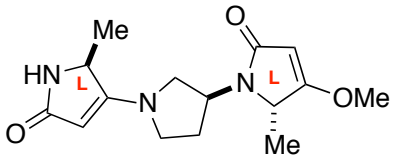
Figure 6.9. continued.

Our Cα - Cβ vector matching program run on compounds **5**, **6**, and **2a** - **h** in the pyrrolinone-pyrrolidine oligomer series (Table 6.2) revelaed **5** and **3a** in the omegatide series matched very similarly on the ideal secondary structures; this is unsurprising because both these compounds present amino acid side chains corresponding to contiguous amino acids. However, compound **6** that mimics *non-contiguous* amino acids is a much better mimic for secondary structures than contiguous templates **3**, and **5**. Preferred conformations of **6** have RMSDs less than 0.3 Å for *all* secondary structures; i.e. **6** is an excellent universal peptidomimetic. Figure 6.10a shows the best matching of **6** on α-helix *i*-*i*+3 residues (RMSD 0.14 Å, score 5.4). The conformer shown is only 0.83 kcal/mol higher in energy than the lowest energy conformer. Moreover, a preferred conformer of **6** overlays extremely well on sheet-turn-sheet *i*+1-*i*' residues (RMSD 0.08 Å, score 2.5; Figure 6.10b). The conformer is only 0.59 kcal/mol higher in energy than the lowest energy conformer. Procedure for overlays and overlays of preferred conformations of **6** on the remainder of secondary structures are given in Appendix F.

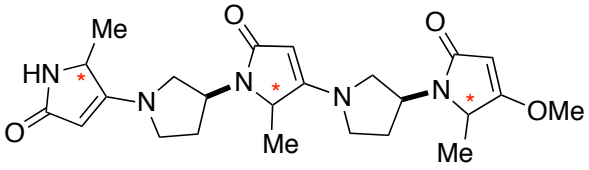
Table 6.2. Evaluation of Preferred Conformers of Mimics **5**, **6** and **2a** - **h** Overlaid on Secondary Structures.



5



6



2a (LLL), 2b (DLL), 2c (LDL), 2d (LLD)
2e (LDD), 2f (DLD), 2g (DDL), 2h (DDD)

com'd	3_{10}-helix		α-helix		π-helix		β-strand		parallel β-sheet		sheet-turn-sheet	
	R[a]	S[b]	R	S	R	S	R	S	R	S	R	S
5	0.55	22.0	0.50	22.1	0.54	25.7	0.46	10.3	0.39	8.7	0.56	8.3
6	0.22	9.9	0.14	5.4	0.14	6.8	0.19	7.5	0.10	4.2	0.08	2.5
2a	0.58	35.1	0.78	33.8	0.63	18.9	0.73	32.2	0.42	23.9	0.46	27.3
2b	0.62	26.2	0.67	24.4	0.67	45.6	0.57	23.6	0.42	22.9	0.34	20.0
2c	0.70	41.5	0.64	31.9	0.54	21.3	0.44	21.0	0.47	25.0	0.34	20.1
2d	0.87	53.1	0.81	49.9	0.83	43.9	0.61	21.5	0.46	16.6	0.36	16.7
2e	0.83	30.8	0.76	31.8	0.64	40.0	0.47	25.6	0.33	21.3	0.36	18.6
2f	0.51	30.5	0.60	26.4	0.64	35.5	0.48	26.1	0.32	10.5	0.34	15.0
2g	0.69	42.9	0.62	33.8	0.59	24.9	0.46	22.5	0.35	18.1	0.33	18.8
2h	0.72	37.2	0.65	30.5	0.57	29.8	0.68	28.8	0.54	22.7	0.44	29.1

[a] R means RMSD (Å). [b] S means score.

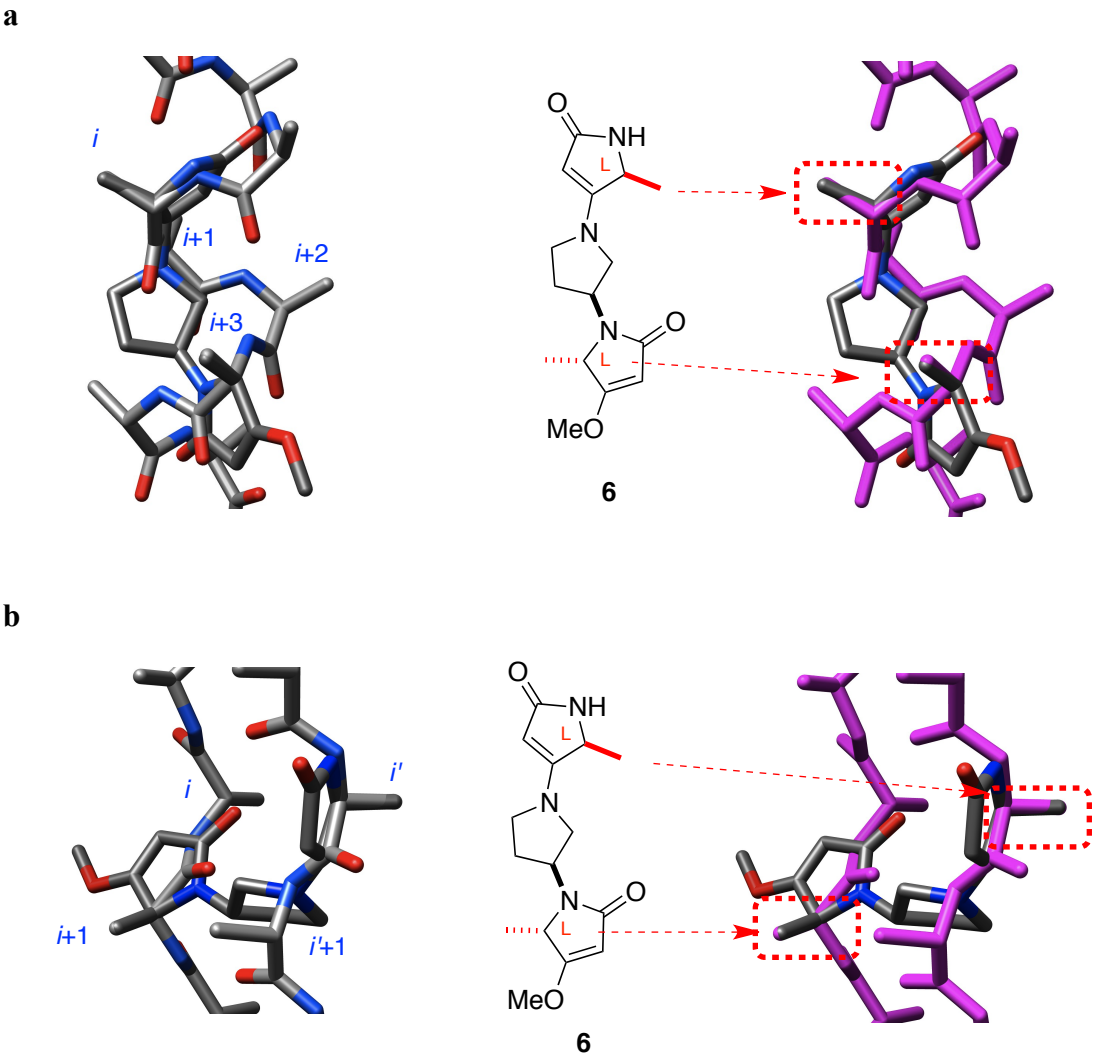


Figure 6.10. Overlays of preferred conformations of **6** on (a) an ideal α-helix *i*-*i*+3 residues; and (b) an ideal parallel β-sheet *i*+1-*i*' residues.

In the structures with three side-chains (pentamers), non-contiguous compound **2** (RMSD 0.32 – 0.87 Å, score 10.5 – 53.1) has preferred secondary structures that overlay better than contiguous compound **1** (RMSD 0.41 – 1.14 Å, score 15.4 – 61.4) on all the featured secondary structures. Analyses for the eight stereomers **2a** - **h** indicate that this type of structure is a better mimic for strands than for helices; almost all RMSDs for

strands including β-strand, parallel β-sheet, and sheet-turn-sheet are less than 0.5 Å (blue highlights in Table 6.2). Stereomers **2f** (D,L,L) and **2g** (L,D,D) also show good fits for helices compared with other stereomers. Figure 6.11a shows the best match (**2f**, RMSD 0.32 Å, score 10.5, ΔG^{o}= 1.28 kcal/mol) in this series on parallel β-sheet. Interestingly, the three side chains of **2a** can overlay well with an ideal sheet-turn-sheet (RMSD 0.46 Å; score= 27.3, ΔG^{o}= 1.93 kcal/mol, Figure 6.11b), and we found one example of protein-protein interaction between monomers in the Rad52 undecamer where **2a** matched three side-chains with an RMSD of only 0.14 Å (Figure 6.11c).

a

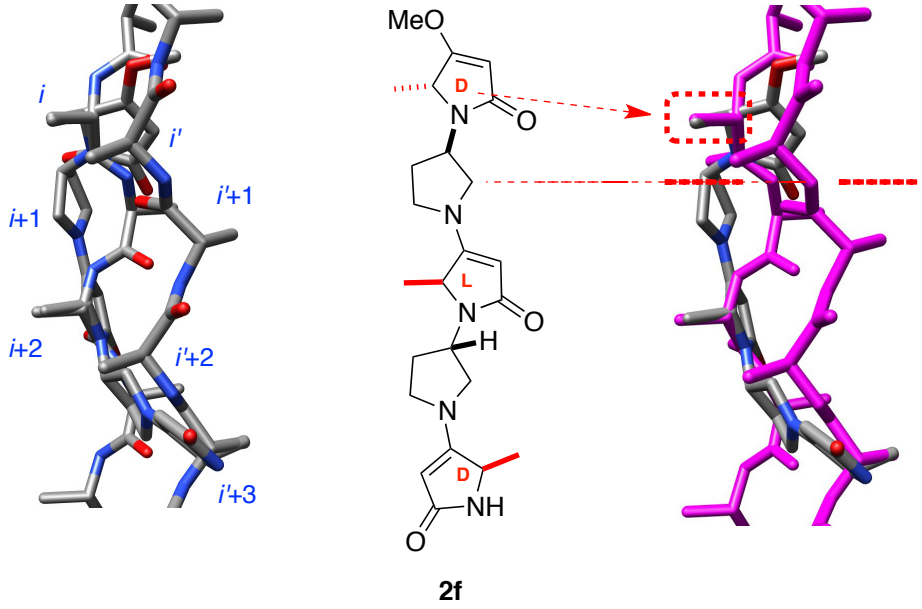
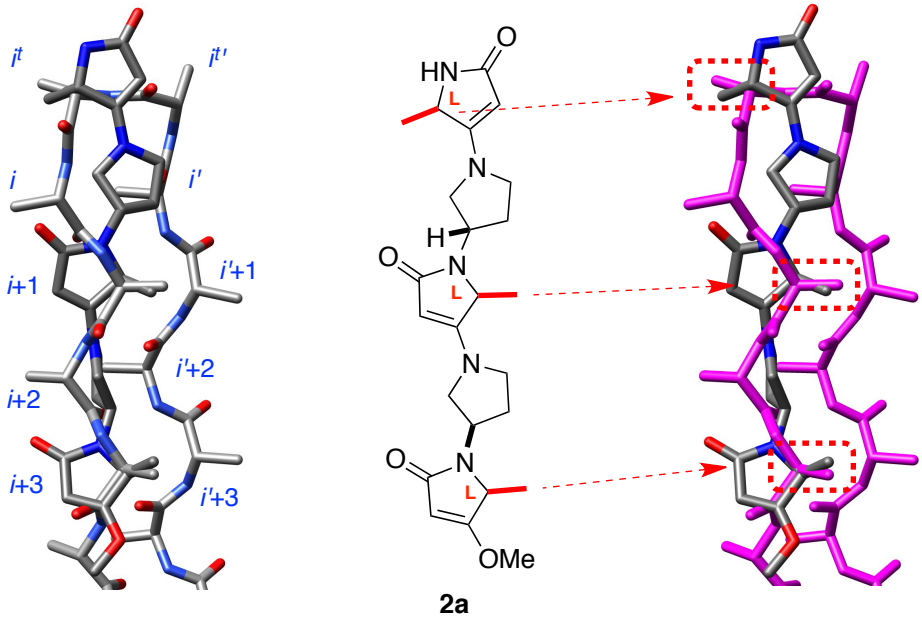


Figure 6.11. (a) Overlay of **2f** on parallel β-sheet motif. (b) Overlay of **2a** on sheet-turn-sheet motif. (c) Matching of **2a** with a monomer of Rad52.

b



c

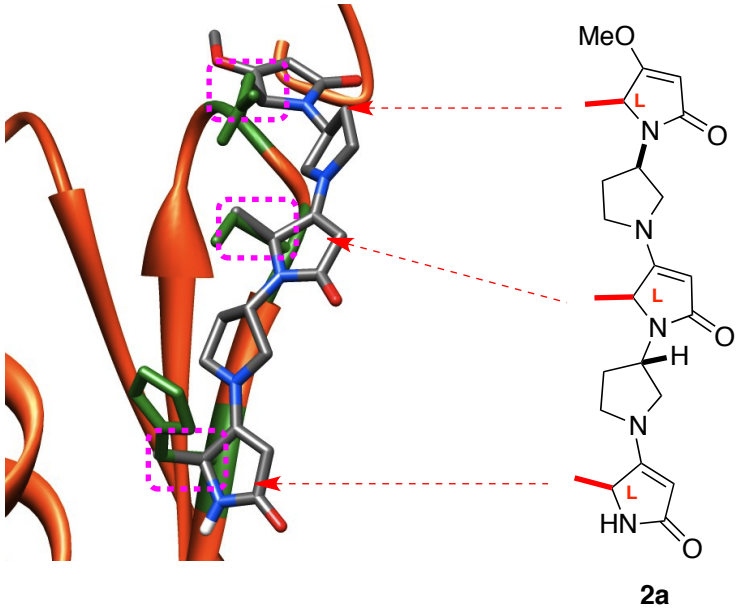


Figure 6.11. continued.

6.5 Conclusion

Secondary structures mimics have traditionally been studied because of their relevance to drug design. Our group suggested four structural criteria for design of effective minimalist mimics of secondary structures: (i) facile syntheses with most amino acid side-chains; (ii) kinetically and thermodynamically accessible conformations for induced fit; (iii) only moderate loss of entropy on docking; and (iv) appropriate Cα - Cβ coordinates of an accessible conformation of the mimic matching to the secondary structures. However, prior to the work described here, there was not a clearly defined computational method to evaluate preferred orientations of Cα - Cβ coordinates via molecular modeling. We developed the Cα - Cβ vector matching program to do this, and applied it to new templates including ones with three amino acid side-chains. Omegatides **1**, and pyrrolinone-pyrrolidine oligomers **2** showed better matches for strand structures than for helices. Interestingly, non-contiguous structures **2** gave better fits for all secondary structures than contiguous structures **1**. Increased degrees of freedom from the pyrrolidine linker in **2** give more flexibility and more appropriate spacing for good fits to secondary structures.

At this stage of the research we realized a much higher goal is possible: Cα - Cβ vector matching to evaluate goodness of fit of small molecule preferred conformations on any combination of three interface side-chains in all structurally characterized protein-protein interactions (PPI). This motivated us to develop programs to do this; specifically, we created a matching program for selected structurally characterized PPIs (eKO), and a PDB mining program (EKO). The next chapter describes how these programs were developed and applied.

CHAPTER VII

THE MATCHING ALGORITHM

"EXPLORING KEY ORIENTATIONS (EKO)"

7.1 Introduction

Protein-protein interactions (PPIs) are involved in most of the essential processes that occur in living organisms, so there have been many attempts to discover small "druggable" molecules that can perturb these.[129] Discovery strategies in this field include high-throughput screening (HTS), fragment-based methods, and computer-aided drug discovery.[19] HTS typically features hundreds of thousands of compounds, but the "hit rates" obtained for PPIs are far less than for traditional targets like enzymes and G-coupled protein receptors.[129,130] Fragment-based methods use small libraries of molecular fragments screened for binding against a target. However, the fragments almost invariably have low binding affinities, so it is necessary to use high compound concentrations and these may give spurious results in biochemical assays.

Peptide-centered approaches are distinct from those featuring small molecules. Peptides are powerful tools to target proteins, but peptide leads have poor pharmacokinetics properties including low cell membrane permeability, and metabolic degradation. Consequently, we do not discuss them further here.

Computer-aided drug discovery is fast, and requires less resource than other methods. There are at least three approaches in the computer-aided drug discovery methods for PPIs: structure-based (featuring docking), complete ligand-based, and ligand fragment (corresponding to experimental fragment-based methods, e.g. NMR or X-ray). All these methods use protein structure data and/or compound databases to find lead molecules. However, docking approaches for shallow clefts at PP-interfaces are unreliable, as are methods to match the physiochemical properties of small molecules

with those of the protein surfaces. In fact, we hypothesize there are several major problems with this approach in general:

- desirable chemotypes of small molecules for interfering with PPIs have not been defined;
- protein conformations may adjust when a small molecule interacts, and the conformation of the molecule may similarly change;
- selectivity issues; and,
- there is no obvious way to relate the structures of molecules found in HTS libraries to the protein ligands that bind to the protein of interest.

Of course, these problem are accentuated for PPIs that are not structurally characterized.[33]

Here we suggest a set of desirable chemotypes for small molecules that may perturb PPIs. Observers may justifiably question if these chemotypes are ideal, but their acceptance as a working hypothesis means that all the other issues outlined above become solvable for structurally characterized PPIs. The genesis of our thinking arose from aspects of the work already discussed in this thesis, as outlined below.

First, we expanded the Cα - Cβ vector matching idea to include any combination of at least three side-chains on one protein at a PPI-interface *regardless of the secondary structures involved.* Justification for the importance of this idea is as follows. The majority (over about 80 %) of hot-spots for PPIs feature only side-chain interactions.[7] Hot-spots are small sets of cooperative contact residues that complement similar ones on the other interface and account for much of the binding energy.[121] There is no method, computational or otherwise, that *definitively* shows which residues combine to give hot-spots, so it is impossible to reliably identify hot-spots, particularly for PPIs that have not been studied in biophysical experiments. All hot-spots occur at PPI interfaces, so instead of targeting hot-spots, we set out to find *interface-mimics* instead, and present researchers with the opportunities to decide whether or not to follow the ideas generated.

We supposed that users of our approach would routinely consult the literature regarding the hit PPIs to see what residues have been implicated in hot-spots (if any). In any case, compounds that perturb an interface, but not at hot-spots, are *perfectly valid* candidates to modulate protein-protein interactions via allosteric effects. In summary, identification of *interface-mimics* is likely to give good pharmacological probes. Conversely, nobody knows how to definitively identify hot-spots, and it would be unnecessarily exclusive to focus on *hot-spot mimics* alone.

Much research on molecules closely related to peptides has focused on mimicry of secondary structures as a way to perturb PPIs. However, hot-spots in protein-protein interfaces are not limited to secondary structures; they may take side-chains from several secondary structure types or from poorly defined folds. For instance, the putative hot-spot residues of epidermal growth factor (EGF, a growth hormone that interacts with epidermal growth factor receptor 1, EGFR1;ErbB-1;HER1) are Tyr13, Leu15, Met21, Ile23, Leu26, Arg41, and Leu47 (Figure 7.1).[131] These hot-spot residues are not confined to defined secondary structures. Finally, it is reasonable to question the justification for broadly analyzing all side chains at PPI-interfaces rather than restricting considerations to hot-spots. Hot-spots are small sets of cooperative contact residues that complement similar ones on the other interface and account for much of the binding energy.[25] There is no method, computational or otherwise, that *definitively* shows which residues combine to give hot-spots. Thus we set out to find *interface-mimics*, and present users with the opportunity to decide whether or not to follow this virtual hit. Users should routinely consult the literature regarding the hit PPIs to see what residues have been implicated in hot-spots (if any). In any case, compounds that perturb an interface, but not at hot-spots, are *perfectly valid* candidates to modulate protein-protein interactions via allosteric effects. In summary, identification of *interface-mimics* is likely to give good pharmacological probes. Conversely, nobody knows how to definitively identify hot-spots, and it would be unnecessarily exclusive to focus on *hot-spot mimics* alone. Therefore, we wanted to expand our focus from secondary structures to whole interfaces in protein-protein interfaces.

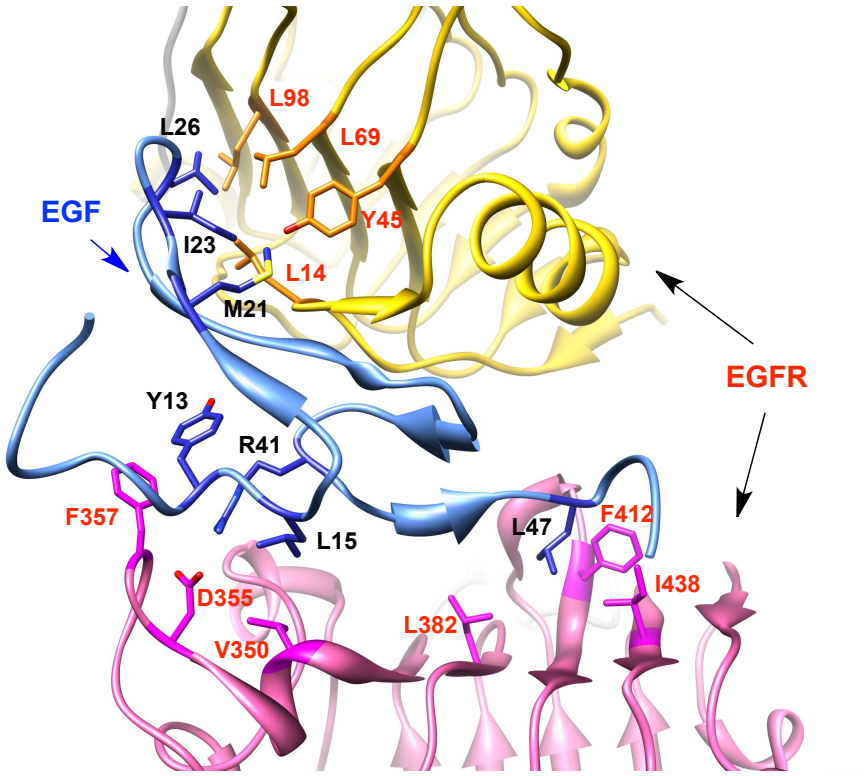


Figure 7.1. Protein interface between EGF and EGFR (PDB 1ivo).[132] Residues shown on EGF are putative hot-spots.

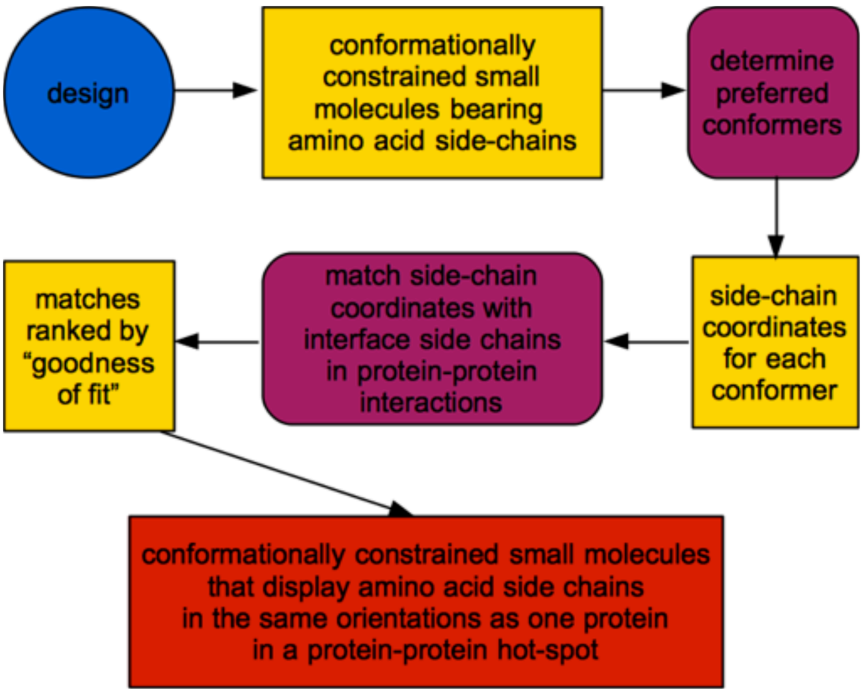
The approach we formulated to find interface mimics is illustrated in Figure 7.2a. Thermodynamically preferred conformations (from QMD) of small molecule scaffolds conforming to our chemotypes are systematically overlaid on any combination of three side-chains at PPI-interfaces, and then scored for "goodness of fit". Just as with our previous work on matching small molecules to secondary structures, this approach is based on Cα - Cβ vectors; Cγ - Cδ orientations and those of more distal bonds in side chains are less important because they are relatively free to rotate. Moreover, as described in Chapter 6, we focus on sets of three side-chains because this is computationally manageable, and simultaneously, we hypothesize that three side-chains are enough to give good affinities and selectivities. Sets of six coordinates each

corresponding to three side-chain Cα - Cβ vectors characterize and differentiate special orientations of side-chains in the preferred conformations (Figure 7.2b). I wish to stress that use of all thermodynamically accessible conformations of molecules (as obtained by QMD) rather than single conformations is important because it reflects the fact that small molecules conforming to our chemotypes do not adopt one single conformation.

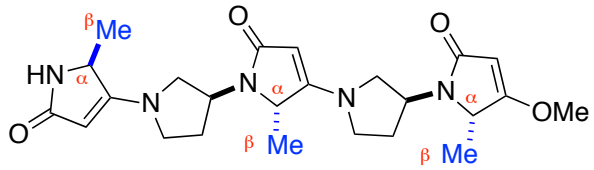
We model only compounds with all three side-chains set as methyl (Ala-Ala-Ala derivative) since our approach only relies on Cα - Cβ coordinates. It is not necessary to mine using full side-chains that correspond to each PPI target because the orientations of the side-chains are largely determined by the Cα - Cβ coordinates of the first side-chain bond in the preferred conformations, and they correspond to these coordinates in the structural data too.

In our method, preferred conformations of the small molecule are simulated in a *featureless* medium of dielectric 80 (corresponding to water). We hypothesize that preferred conformations of a molecule with three methyl side-chains in this medium will be even more preferred for corresponding compounds with fully functionalized side-chains bound to a protein-binding partner. Perturbations to the preferred conformations of the scaffold with other side-chains in the absence of the protein are inconsequential.

a



b



1

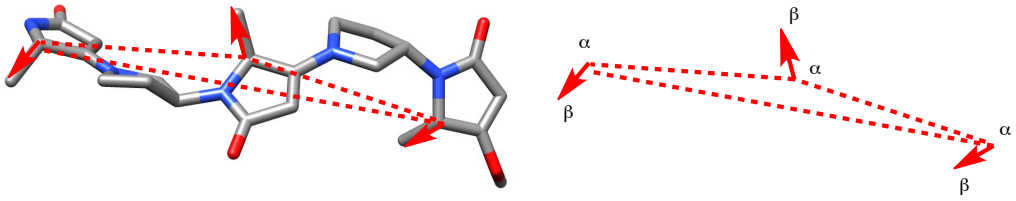


Figure 7.2. (a) EKO matches key side-chains in PPIs with preferred conformations of rationally designed molecule. (b) Six Cα - Cβ coordinates in a small molecule define the side-chains' projection in space.

In the key step of our matching algorithm, all accessible conformations from QMD are paired on all relevant areas of PPI-interfaces. That is, our system screens accessible conformations from our molecule against all combinations of three side-chains at all the distinct interfaces in structurally characterized PPIs. This mining procedure is achieved in one analysis that takes about 6 h on the TAMU supercomputer system.

The issue outlined above about protein and small molecule conformations adjusting in interaction, is circumvented when matching preferred conformations of small molecules to conformations of side chains at structurally characterized PPI-interfaces. This is because it is known that the PPI-interface regions preferentially crystallize in these structures, and the conformations of the small molecule that match them are also preferred. In other words, concerns based on the small molecule and the PPI interface adjusting to each other are insignificant because *the orientations of side chains in both components are known to be favorable.*

Our proposed solution to the selectivity issue outlined above relies on the fact that our molecules are based on any combination of side-chains expressed on the scaffold with any relative stereochemistry. Just as peptides have *different* diverse functions based on the composition and orientations of their side-chains, compounds based on our chemotypes will gain affinity *and* selectivity as a result of the same factors.

In summary, our method is outlined in Figure 7.2a may be regarded as "virtual affinity chromatography" where all structurally characterized PPIs are passed through an affinity column bearing the featured small molecule bait, free to adopt any of its preferred conformations. We think this approach is novel. It is unique to mine preferred conformations of a small molecule on PPI-interfaces. The closest method may be Bartlett's CAVEAT program. This can use Cα - Cβ vectors to pair preferred conformations of small molecules to a structurally characterized substrate-enzyme complex to facilitate design of inhibitors.[133] The key differences are that CAVEAT: (i) only considers one preferred conformation of the small molecule rather than the holistic approach that QMD provides; (ii) matches those single preferred conformations of small

molecules on enzyme substrate fragments that bind in active sites (not on protein side-chains at PPI-interfaces); and, (iii) has not been used for mining huge structural databases featuring proteins. DOCK is another well known algorithm that is similar to CAVEAT.[134] This operates by calculation low energy binding of interactions of small molecules in active sites of a macromolecule. DOCK does not focus of Cα - Cβ coordinates, it does not use data from protein side-chains at PPI-interfaces, and it has never been used for mining huge structural databases of PPIs.

7.2 Development of The EKO Algorithm

To achieve the goals we had identified above we needed a mining algorithm. The one that is now developed we call Exploring Key Orientations (EKO). It would not have been possible to create EKO without considerable help from experts in the writing and implementation of computational algorithms; consequently, we are indebted to Dr Thomas Ioeger of the Computer Science Department, for writing a script on which EKO is based, and for modifying it according to the issues we encountered when we attempted to use it. Additionally, Dr Lisa M. Perez of The Molecular Simulations Laboratory at Texas A&M University played a pivotal role: she developed a script for running QMD, interfaced it with the algorithm, and refined the integrated package so that it ran on the supercomputer resource at TAMU.

Recall that conformational analyses with QMD typically generate 600 low energy structures. The first step in the implementation of the EKO algorithm is to input structural data based on PPIs. EKO does not use raw data from the PDB. After some searching, it was decided to base EKO on the protein database generated by "3D complex".[135] 3D complex covers all protein structures that were released before 2008. Several attributes of this feature of 3D complex are ideal for its application in EKO, specifically it:

- has none of the redundancy and problematic PDB structures mentioned above; and,
- identifies contacts between the chains.

The 3D complex database encompasses over 53,000 chain interactions in 15,736 structures. Not all the side-chains are pertinent to PPI-interfaces, so a user defined filter is applied to focus EKO on the most pertinent side-chains. For a side-chain to be considered it must be within X Å of the other protein chain; typically, this distance "X" is set at 4 Å.

Having applied the filter, EKO enumerates all combinations of three amino acids in the protein-protein interface. She superposes each conformation of the molecule onto each combination of interface residues. This is based on optimal superposition of three pairs of Cα - Cβ vectors from each conformation onto a triplet of amino acid side-chains in a protein chain.

Each superposition is scored for goodness of fit, and the best candidate is returned. The scoring system is the same as the ones described in Chapter 6 based on RMSD and "score". When we applied EKO with 368 conformations from compound **1** for 15,736 different crystal structures from 3D complex, she found 186 hits corresponding to conformations of the small molecules that overlay on interface residues with RMSDs of 0.33 Å or less.

To overcome this issue that we can screen only protein structures that were released before 2008, we developed a sister algorithm, "eKO" as a matching program for each crystal structure on the same principles. eKO can; (i) match any good crystal structure, including those released after 2008; (ii) use all the conformations of the molecule without clustering; and, (iii) match for less CPU expenditure because only select PPI structures are considered. The 3D database that EKO feeds on does not include NMR structures, but eKO can evaluate these from the PDB. eKO is not intended for mining huge databases. The remainder of this chapter describes applications of EKO and eKO for compound **1**. The dimer interface of HIV-1 protease was one of the several PPIs implicated from this analysis, so I took samples of **1** (prepared by Dr Arjun Raghuraman) and tested them to obtain experimental data to validate the approach.

7.3 Application of EKO to Pyrrolinone-pyrrolidine Oligomers 1

QMD analysis of **1** generated 490 conformations in 166 families within 3.0 kcal/mol of the global minimum; members of each family cluster on the Cα - Cβ coordinates to within 0.5 Å RMSD or less. A computational subroutine was used to select representative conformers from these families according to the specifications outlined above; consequently, 368 conformers for molecule **1** were chosen for the next step. EKO sampled these 368 conformers on 1,340,343 PP-interface residues that remained after the distance filter (4 Å) were applied to the 53,328 protein chain interactions in the 15,736 structures covered by the 3D database. EKO scored each superposition, and then generated lists of PPIs that best match **1** based on RMSD of the small molecules preferred conformations. Specifically, 186 hits within RMSD 0.33 Å were found in this process (Figure 7.3).

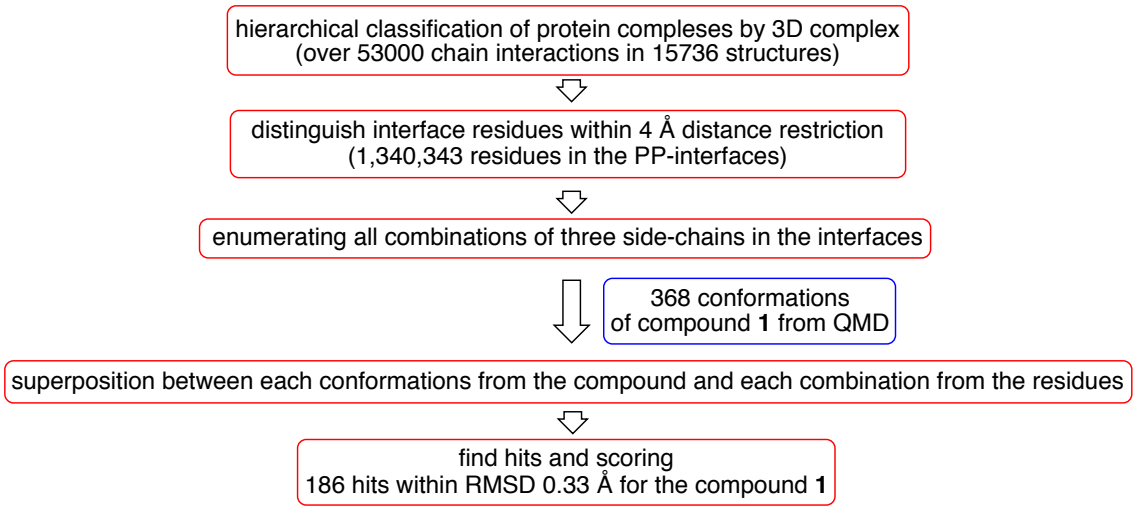
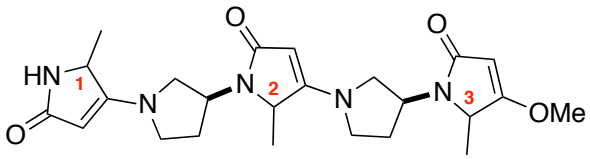


Figure 7.3. Mining preferred conformations of **1** on PP-interface side-chains using EKO.

EKO was similarly applied for all eight stereomers of compound **1**. Table 7.1 summarizes matches for all the stereoisomers. To keep the numbers manageable, an RMSD 0.30 Å or less was chosen for all stereoisomers except LLL for which 0.33 Å was used; our thinking at the time was that we wanted to consider more hits for the "natural" stereoisomer.

EKO found 183 hits for the LLL-isomer **1** corresponding to 58 different proteins. Redundancies occurred when different conformations overlay on the same protein with different RMSDs below the cutoff, and when the same conformations of the molecule overlay on different crystal structures of the same proteins. In this series of stereomers, **3** (LDL) gave most hits on the largest set of different proteins.

Table 7.1. The PDB Mining Results for Eight Stereomers of Compound **1**.

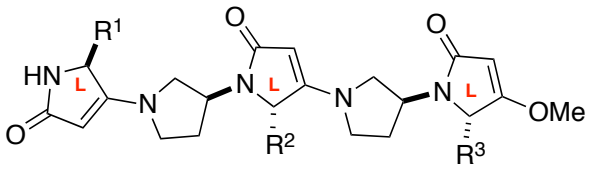


compounds	stereomers (1-2-3)	# of results[a]	# of unique hits
1[b]	LLL	183	58
2	DLL	197	76
3	LDL	304	130
4	LLD	117	37
5	LDD	55	22
6	DDL	104	50
7	DLD	57	29
8	DDD	36	15

[a] Results within RMSD 0.3 Å. [b] Results within RMSD 0.33 Å.

Table 7.2 presents mining results for stereomer **1** (LLL) within RMSD 0.33 Å. The best match was found for a monomer-monomer interface of a human protein called Rad52 (RMSD of only 0.14). Analysis of the interface regions that matched showed compound **1** has a bias towards β-sheet regions, including strand, parallel β-sheet, and sheet-turn-sheet structures. This observation correlates with the predictions from Cα - Cβ vector matching on secondary structures as described in Chapter 6. However, most of the matches corresponded to undefined secondary structures or ones in which the mimic spanned two or more types. This is consistent with the hypothesis we presented earlier that there is no absolute correspondence between secondary structure and interface (or hot-spot) mimics.

Table 7.2. Mining Analyses for Stereomer **1 (**LLL).



1

entry	PDB	proteins	RMSD (Å)	score	residues (R^1-R^2-R^3)	secondary structure
1	1kn0	Rad52	0.14	7.0	H121-S119-D117	strand-loop
2	1n2c	nitrogenase	0.19	11.7	K145-D76-S257	
3	1g0o	trihydoxynaphtalene reductase	0.23	12.3	P173-H122-V126	
4	1j3u	aspartase	0.23	13.8	V236-T234-V232	loop
5	1gl7	TrwB	0.23	15.6	T352-D349-S346	helix-loop
6	1six	trypsin-ecotin	0.24	11.3	Me5-T83-L52	

Table 7.2. continued.

entry	PDB	proteins	RMSD (Å)	score	residues (R^1-R^2-R^3)	secondary structure
7	3pcb	3,4-PCD[a]	0.24	15.7	Q177-175-K173	
8	1fcj	*O*-acetylserine sulfinydrylase	0.24	17.0	L268-S301-E303	
9	2f4f	IS200 transposase	0.25	14.5	H60-V18-V107	β-sheet
10	1mtp	serpin	0.26	11.8	Y200-T210-A218	β-sheet
11	1eef	heat-labile enterotoxin	0.26	13.5	T47-I39-E29	
12	1b99	NDP[b]	0.26	16.1	V46-Q44-L42	β-sheet
13	2bk6	ferroxidase	0.26	17.6	L65-P71-S73	
14	1w0k	Bovine F1-ATPase	0.27	16.3	E84-I34-D36	
15	1thz	AICAR Tfase[c]	0.28	17.8	A218-L220-T222	β-sheet
16	3gpd	GAPDH[d]	0.28	12.3	T228-M230-F232	β-sheet
17	1wwh	MPPN domain	0.28	14.6	F30-N60-T58	
18	1d2r	TrpRs[e]	0.28	15.8	L300-I79-D41	
19	2awd	LacC (togatose-6-phosphate kinase)	0.28	16.2	D91-I15-V33	β-sheet
20	1iql	human coagulation factor Xa	0.28	17.6	W114-E111-L109	helix-loop
21	2dys	cytochrome C oxidase	0.28	18.9	P176-L160-I199	
22	1de4	HFE/β2m	0.28	19.1	D96-Y10-R12	
23	1hpv	HIV-1 protease	0.29	11.5	L97-C95-I93	β-sheet
24	1oux	lectin LecB	0.3	14.7	R13-P112-L87	β-sheet
25	1lem	lentil lectin	0.3	14.1	S7-S5-T3	β-sheet
26	1m3y	capsid protein	0.3	17.0	Q120-R122-E49	β-sheet
27	2nvx	RNA polymerase II	0.3	17.7	R1159-Q1193-L1191	β-sheet

Table 7.2. continued.

entry	PDB	proteins	RMSD (Å)	score	residues (R^1-R^2-R^3)	secondary structure
28	3fru	neonatal Fc receptor	0.3	18.5	F22-L65-Y63	β-sheet
29	1mdl	mandelate racemase	0.3	18.6	P77-A79-S82	
30	1af6	maltoporin	0.3	19.0	S9-I11-W13	β-sheet
31	1n1q	Dps protein	0.3	20.7	A73-P71-L65	
32	1rav	recombinant avidin	0.3	21.0	K93-M95-L97	β-sheet
33	1s5c	cholera holotoxin	0.31	13.9	T47-I39-E29	
34	1ukj	L-methionine-lyase	0.31	14.5	Q34-S248-A87	
35	1pmo	glutamate decarboxylase (GadB)	0.31	16.1	S318-P320-Q323	
36	1ko0	diaminopimelate decarboxylase	0.31	15.9	A366-V287-Q289	
37	1go3	RPB4/RPB7	0.31	20.1	E88-E86-L84	β-sheet
38	1i7t	class I MHC HLA A2	0.31	20.1	Q32-V25-I23	β-sheet
39	1hng	T lymphocyte antigen glycoprotein	0.31	20.5	P175-N173-V171	
40	1nnu	enoyl acyl carrier reductase	0.31	20.6	A216-I214-I167	
41	1y14	Rpb4/7	0.31	21.4	V86-D88-T90	β-sheet
42	2gyy	ASADH[f]	0.32	12.5	L206-I246-V248	
43	1t3n	DNA polymerase	0.32	15.1	L361-R317-C383	
44	1hle	leucocyte elastase inhibitor	0.32	16.4	R27-L25-L23	β-sheet
45	1rs8	NOS[g]	0.32	17.0	A358-D355-L343	
46	1huu	HU protein	0.32	17.2	F50-P77-K75	
47	2ltn	lectin	0.32	17.6	V116-I139-N148	

Table 7.2. continued.

entry	PDB	proteins	RMSD (Å)	score	residues (R^1-R^2-R^3)	secondary structure
48	1xs4	dCTP deaminase	0.32	19.3	L88-L90-V138	β-sheet
49	2mpr	maltoporin	0.32	20.6	W13-I11-S9	β-sheet
50	1shs	heat-shock protein	0.33	14.8	V128-I57-I47	
51	1m5h	formyltransferase	0.33	14.2	I201-Y245-A243	
52	1lob	isolectin I	0.33	19.3	I87-F85-F83	β-sheet
53	1vgr	formyl-CoA transferase	0.33	19.4	C144-A150-R252	
54	1qqw	catalase	0.33	20.3	P341-A418-Q415	
55	1ohh	F1-ATPase	0.33	20.7	E84-I34-D36	
56	2f0c	receptor binding protein	0.33	21.8	T187-T185-R183	β-sheet
57	1ama	aspartate aminotransferase	0.33	22.7	V35-A37-K250	
58	1rj7	EDA-A1	0.33	23.7	A121-V123-L125	β-sheet

[a] 3,4-PCD: protococatechute 3,4-dioxygenase. [b] NDP: nucleoside diphosphate kinase. [c] AICAR Tfase: avian aminoimidazole-4-carboxamide ribonucleotide transformylase. [d] GAPDH: D-glyceraldehyde 3-phosphate dehydrogenase. [e] TrpRs:tryptophanyl-tRNA-synhtetase. [f] ASADH: aspartate semialdehyde dehydrogenase. [g] NOS: nitric-oxide synthase.

Data in Table 7.2 suggests derivatives of compound **1** with appropriate side-chains are candidates for disruption of several PPIs of biomedicine importance. These potential targets include: heat-labile enterotoxin/cholera toxin, aminoimidazole-4-carboxamide ribonucleotide transforylase (AICAR Tfase), D-glyceraldehyde 3-phosphate dehydrogenase (GAPDH), HIV-1 protease, and nitric-oxide synthase (NOS).

Heat-labile enterotoxin and cholera toxin are directly associated with causing cholera and related enteropathies in human and domestic animals.[136] AICAR Tfase catalyzes the last two steps in purine biosynthesis, which is a pathway that becomes activated in cancer cell types.[137,138] GAPDH mediates oxidative phosphorylation of the aldehyde in the glycolytic pathway and implicates in apoptosis of cells associated with several neurodegenerative diseases. [139,140] HIV-1 protease is essential for the viral infectivity and propagation of AIDS.[141] NOS catalyzes biosynthesis of NO and concomitant transformation of L-arginine to L-citrulline; it is expressed in various human disease states including infection (e.g. HIV and Malaria), chronic inflammation, neurodegeneration, and cancer.[142] The next section describes HIV-1 protease as a target for disruption of PPIs in detail. Closer consideration of the other potential targets is included in Appendix G.

7.4 Dimerization Inhibitors for HIV-1 Protease

HIV-1 protease is essential for the viral infectivity, and propagation of AIDS.[141] Combination therapies including HIV-1 protease inhibitors are front-line strategies for slowing the progression of this disease.[143] Nevertheless, this approach is imperfect; strains of HIV resistant to combination therapies have emerged, and there are several undesirable side-effects.[144]

HIV-1 protease exists as a stable homodimer for which the Gibbs energy of stabilization has been estimated to be *ca* 14.5 kcal/mol at 25 ^{o}C (pH 5), corresponding to a dissociation constants about 3.4 nM at 37 ^{o}C.[145,146] Dimers of HIV-1 protease form the active site as illustrated in Figure 7.4, and without which the enzyme cannot be active; it is “an obligatory dimer”.

The dimerization interface in HIV-1 protease is mainly formed by four stranded β-sheets on the *N*- and *C*- termini. Specially, Cys95-The96-Lue97-Asn98-Phe99 and Pro1-Ile3-Leu5 account for about 75 % of the total binding energy.[147] Mutation of Cys95 to Ala has little impact on the protease activity, and presumably on the dimerization energy.[148] This is fortunate because methyl side-chains can be used in

place of $–CH_2SH$ groups, so peptidomimetics for the region are more easily made and manipulated.

Peptides have been designed to perturb the dimerization interface of HIV-1 protease. These were based on the *C*- terminus,[146] the *C*- and *N*- termini separately,[149-151] the *C*- and *N*- termini linked by a hydrophobic chain, [152-154] or *C*- and *N*- terminal peptides linked through a side-chain.[155] These peptide-based compounds are unlikely to be cell permeable, and we can find no evidence in the literature that indicates otherwise. This problem has motivated some researchers to combine these types of peptides with HIV-TAT.[156]

A few peptidomimetics of "dimer-disrupting" peptides for HIV-1 protease have been prepared. Chmielewski's group synthesized a cross-linked peptoid scaffold for the *C*- and *N*- termini; these gave low micromolar IC_{50} values in fluorescence-based inhibition assays for HIV-1 protease.[157] A few peptidic compounds were prepared by incorporating Bartlett's @tide residues and gave compounds with K_d values of about 400 nM.[158] It has also emerged that some non-peptidic compounds that inhibit the enzymes activity by binding the active site can also act as dimerization inhibitors by binding at the protein-protein interface.[159]

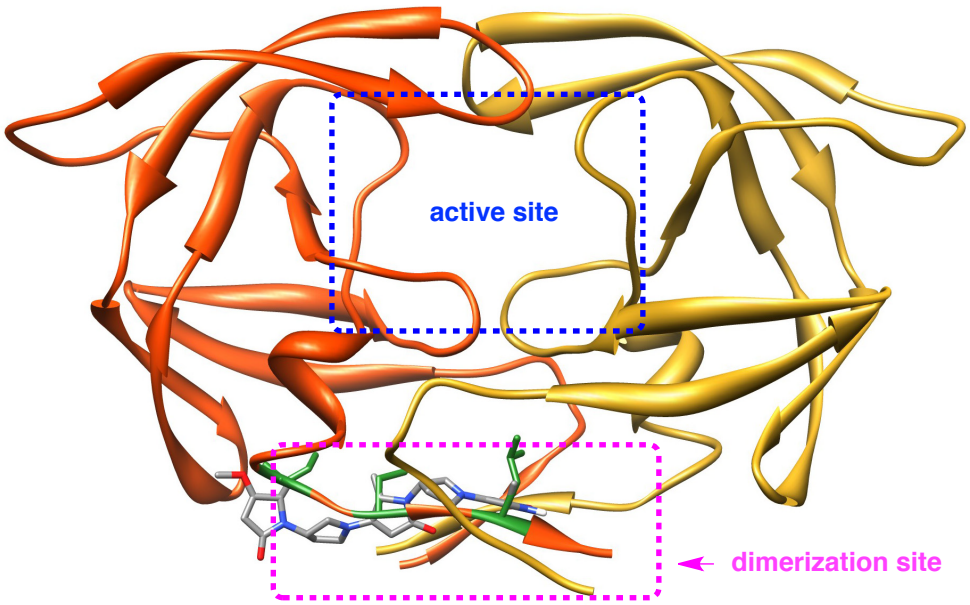


Figure 7.4. Structure of HIV-1 protease from 1hpv.

Mining of the database using EKO for stereomer **1** gave six different results for HIV-1 protease within RMSD 0.33 Å (1hpv, 1gno, 1ytg, 2fde, 1mtb, and 9hvp); in all these cases the *C*- terminal interface residues Ile93-Cys95-Leu97 were implicated. These residues correspond to the region of the HIV-1 protease dimer where the *C*- and *N*-termini interact to form a four-strand sheet network (Figure 7.5). Cys95 and Leu97 correspond to a hot-spot region that was suggested based on thermodynamic studies.[147] The best match (from 1hpv) has an RMSD 0.29 Å and score 11.5, and represents a conformer of **1** that is only 1.24 kcal/mol over the lowest energy conformation located.

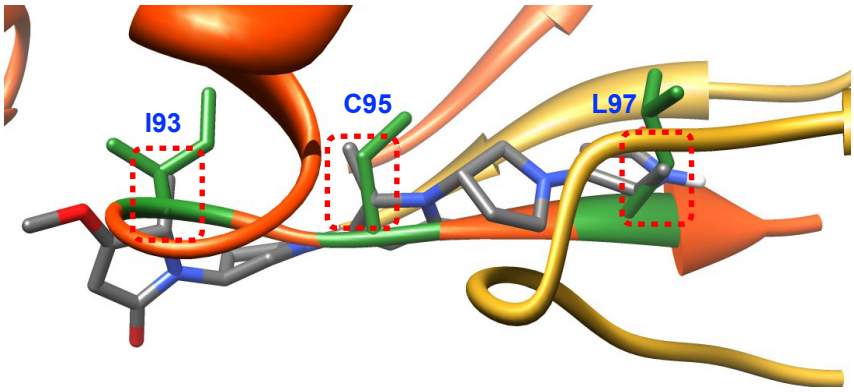
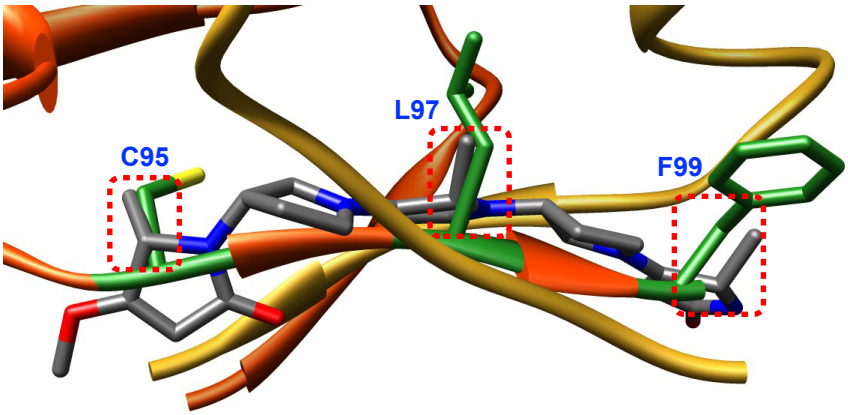


Figure 7.5. Overlay identified from EKO.

We applied eKO to the HIV-1 protease crystal structure 1hpv[160] to find more possible matching regions sets for compounds **1**. This revealed stereomer **1** also overlaid Cys95-Leu97-Phe99, but with a higher RMSD 0.46 Å and score 19.8 (ΔG^o = 1.38 kcal/mol, Figure 7.6a). That residue set exactly corresponds to the hot-spot region on the *C*- terminus. Moreover, we found *N*-terminal matches; specifically, preferred conformers of the template overlaid with Leu5-Ile3-Pro1 (RMSD (0.64 Å and score 22.8; Figure 7.6b). All these three residues are implicated in a hot-spot region.[147] Similar results were obtained from eKO-analyses of other crystal structures (1gno, 1ytg, 2fde, 1mtb, and 9hvp).

a



b

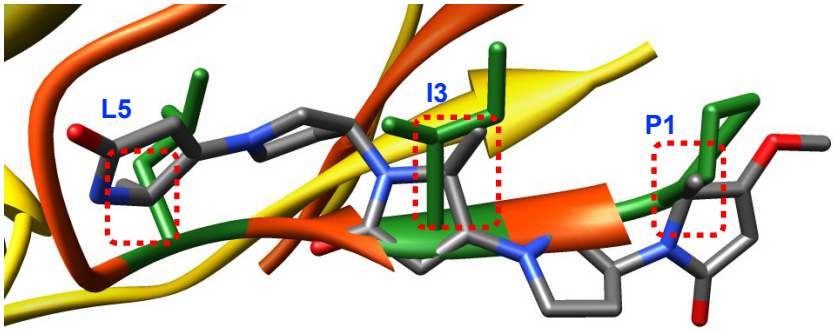


Figure 7.6. Overlays identified from eKO.

Our ability to test compounds based on these analyses were limited because the structures are relatively difficult to make. Dr Arjun Raghuraman could only prepare two deprotected putative analogs of the HIV-1 protease *C*-terminus in the time available. However, he also provided two protected forms, and several intermediates, and I was able to combine two of these intermediates to give a bridged compound with 2 x 2 side chains that we predicted might impact the *C*- and *N*-termini in the dimerization region.

7.5 Bioassays to Validate Disruption of HIV-1 Protease Dimerization

The *C*-protected (**1a** and **1b**) and unprotected (**1a'** and **1b'**) compounds in Figure 7.7b were prepared for an inhibition assay. Intermediates (**9a**, **9a'**, **10a**, **10a'**, and **11a'**) and the bivalent mimic (**12**) in Figure 7.7a and c were also tested.

These compounds have side chains corresponding to the HIV-1 dimerization interface, except that the cysteine side-chain (corresponding to Cys95) was replaced by Ala. Previous studies have shown HIV-1 protease mutants wherein Cys95 was replaced with Ala have almost the same K_d for the dimer dissociation,[148] hence we used Ala instead of Cys95 in syntheses.

a

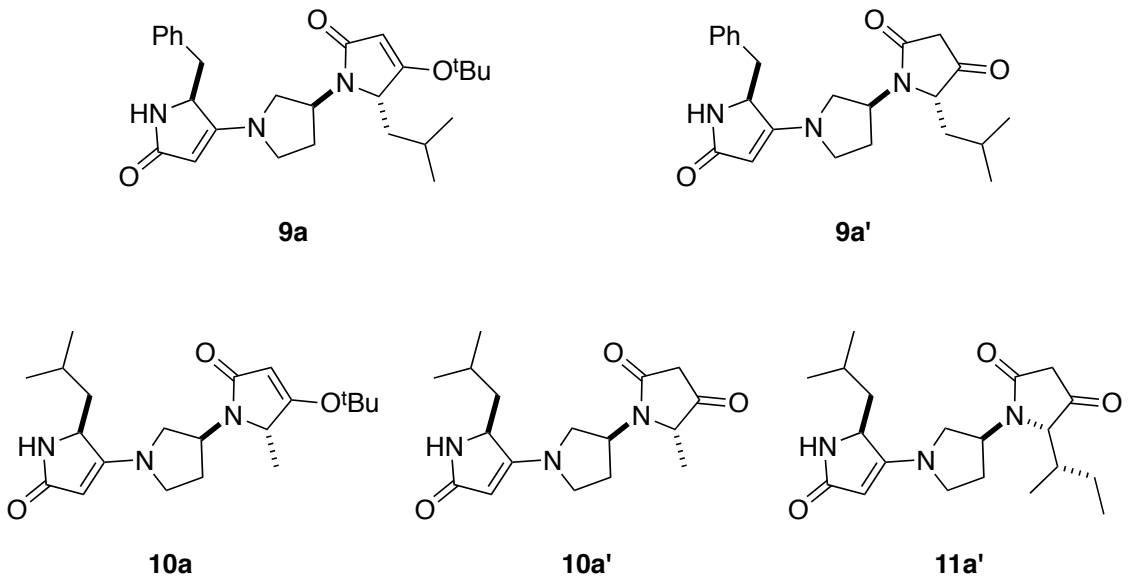
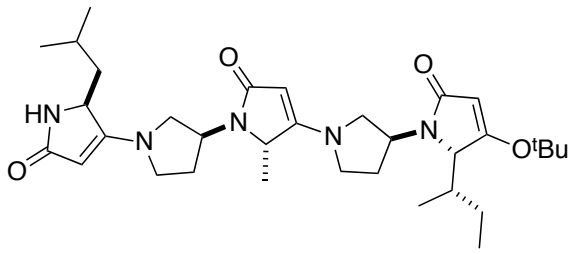
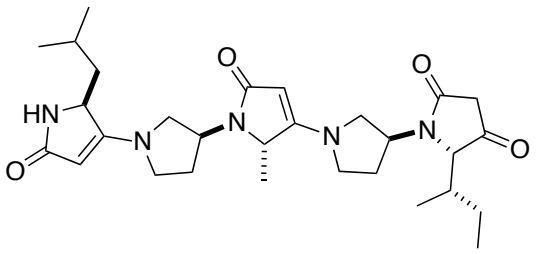


Figure 7.7. Compounds for HIV-1 inhibition assay. Prime on labels indicates the *C*-unprotected forms.

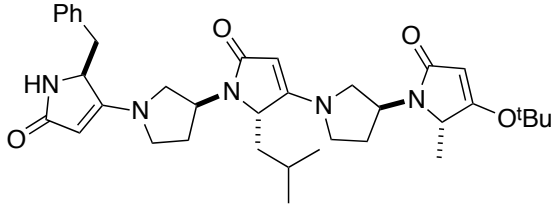
b



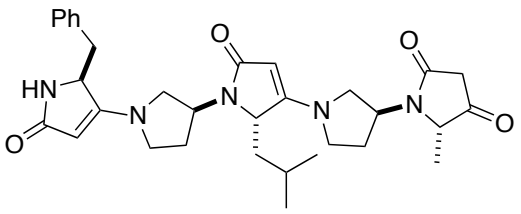
1a



1a'

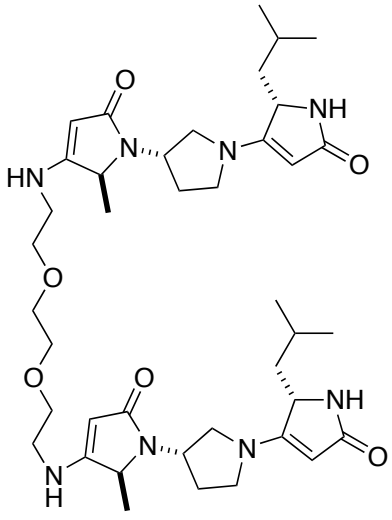


1b



1b'

c



12

Figure 7.7. continued.

The HIV-1 protease inhibitory activities of the compounds were determined by a FRET method.[156,157,161] HIV-1 protease (Q7K) was kindly provided by Dr Celia Schiffer at University of Massachusetts Medical School. HIV-1 protease stock solution was diluted with assay buffer (0.1 M sodium acetate, 1.0 M sodium chloride, 1.0 mM EDTA, 1.0 mM DTT, 10% DMSO, and 1.0 mg/mL BSA, pH 4.7).[156] All inhibitors were dissolved in DMSO, and diluted to appropriate concentrations with deionized water such that the maximum concentration of DMSO in the buffer was 8.5 %. EDANS/DABCYL-based FRET peptide substrate (Ex/Em=340/490 nm) solution in SensoLyte® 490 HIV Protease Assay Kit (Cat. # 71127) and HiLyte Fluor™ 488/QXL™ 520-based FRET peptide substrate (Ex/Em=490/520 nm) solution in SensoLyte® 520 HIV Protease Assay Kit (Cat. # 71147) were purchased from Anaspec. We needed to use these two different substrates because the compounds without *C*-protection (**1a'**, **1b'**, **9a'**, **10a'**, and **11a'**) have weak fluorescence that interferes with that from the EDANS/DABCYL-based FRET peptide substrate (Ex/Em=340/490 nm) so an alternative substrate was used. Concentrations of the substrates are proprietary information of Anaspec, hence we followed a protocol for the substrate preparation in the assay kit.

For the determination of IC_{50} values, HIV-1 protease (40 μL, 10.2 nM final concentration) and inhibitors (10 μL) were incubated for 15 min at 25 ^{o}C. Substrate solutions (50 μL) in the buffer were added into the incubated solution to initiate the reaction. EDANS/DABCYL-based FRET peptide substrate was used for *C*-protected inhibitors, and HiLyte Fluor™ 488/QXL™ 520-based FRET peptide substrate was used for deprotected inhibitors. The total assay volume was 100 μL. Fluorescence was monitored for 5 min at 30 ^{o}C in a fluorescence microplate reader (BioTek) at Ex/Em=340/490 nm for *C*-protected inhibitors, and Ex/Em=490/520 nm for deprotected inhibitors. The initial velocities were plotted against log[inhibitor] and a sigmoidal curve was fitted to the data points using Graphpad Prism 5 software to obtain IC_{50} values.

We obtained the best IC_{50} (3.7 ± 0.3 μM) for inhibition of HIV-1 protease from compound **1a'** (Figure 7.8). Other results are summarized in Table 7.3. Overall, the

compounds with three side-chains **1a**, **1a'**, **1b**, and **1b'** gave better inhibition effects than two side-chain compounds **9a**, **9a'**, **10a**, **10a'**, and **11a'**, and *C*-deprotected compounds **1a'** and **1b'** showed better inhibition than protected forms (**1a** and **1b**). Comparison of **10a**, **10a'** and **12** reveals that compound **10a'** did not show any inhibition, but the corresponding bivalent compound **12** showed two times better inhibition than trimer **10a**.

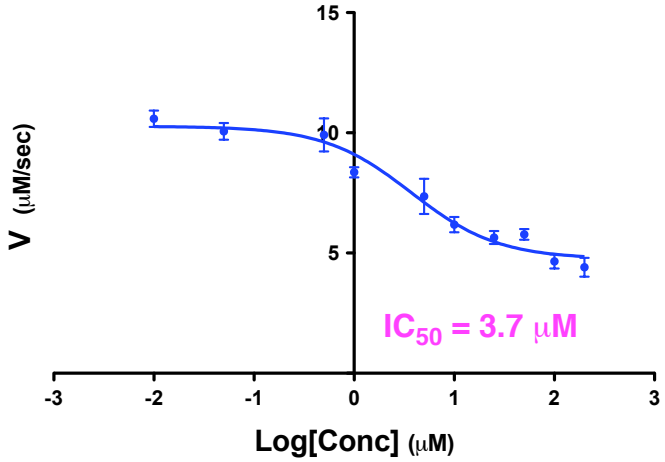


Figure 7.8. Determination of the IC_{50} value for inhibition of HIV-1 protease using **1a'**.

To explore if these compounds inhibit *dimerization* of HIV-1 protease, we carried out a Zhang-Poorman kinetic assay.[146,156] If a compound acts as a dimerization inhibitor, the Zhang-Poorman plot gives a line with a slope similar to that obtained for the uninhibited control but with a different intercept; active-site inhibitors yield different slopes compared with the uninhibited control. For this assay, HIV-1 protease was used at concentrations from 0.6 to 5.1 nM. Substrate solutions were diluted to 1/4 solution from the original solution, and then used for the kinetic assay. HIV-1 protease (40 μL) was incubated with or without an inhibitor (10 μL) at the desired concentration for 15

min at 25 ^{o}C. The diluted substrate solution (50 μL) was added to the incubated solution. Fluorescence was monitored for 15 min at 30 ^{o}C in a fluorescence microplate reader (BioTek) at Ex/Em=340/490 nm for *C*-protected inhibitors, and Ex/Em=490/520 nm for deprotected inhibitors.

Figure 7.9a shows Zhang-Poorman plots for *C*-protected compounds, **1a** and **1b**. Slopes for **1a** (10.4 ± 1.0) and **1b** (9.5 ± 0.6) are similar with one for uninhibited HIV-1 (9.7 ± 0.7) with significantly different y-intercepts (**1a** y-intercepts 2.6, **1b** y-intercepts 1.4, control y-intercepts 0.42); these observations indicate the compounds *are* acting as dimerization inhibitors (see below). The deprotected compounds **1a'** and **1b'** also showed similar patterns as illustrated in Figure 7.9b. Slopes for **1a'** (8.2 ± 0.6) and **1b'** (9.3 ± 1.2) compared with uninhibited HIV-1 (9.0 ± 0.6) have different y-intercepts (**1a'** y-intercepts 0.19, **1b'** y-intercepts 0.73, control y-intercepts 0.013). y-Intercepts of uninhibited HIV-1 have different values between experiments for *C*-protected and deprotected compounds because the substrates used in the assay are different (see above). The results are consistent with **1a**, **1a'**, **1b**, and **1b'** acting as dimerization inhibitors. K_i values calculated from the y-intercepts using the Zhang-Poorman equation are summarized in Table 7.3. We note that it has been reported that inhibition of HIV-1 protease activities by dimerization inhibitors is dependent on the time of pre-incubation with the enzyme and inversely dependent on enzyme concentration; consequently, these factors must be standardized if comparing our data with those from different labs.

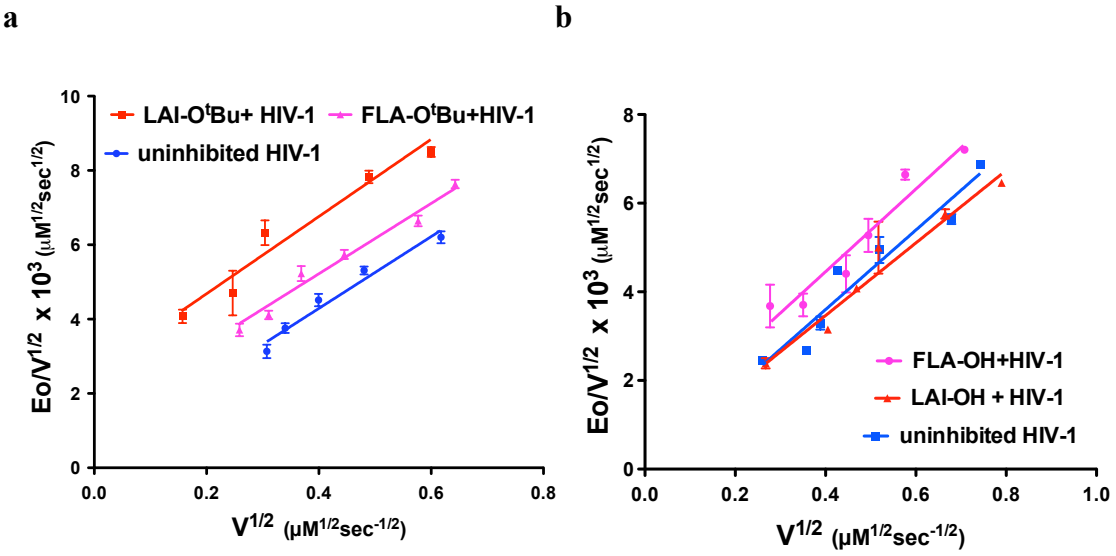


Figure 7.9. Zhang-Poorman analyses for (a) **1a** (LAI-O^{t}Bu), **1b** (FLA-O^{t}Bu), and uninhibited HIV-1 protease; and, (b) **1a'** (LAI-OH), **1b'** (FLA-OH), and uninhibited HIV-1 protease.

Table 7.3. Summary of IC_{50} and K_i.

compounds	sequences	IC_{50} (μM)	K_i (μM)
9a	FL-O^{t}Bu	516.3	
9a'	FL-OH	418.7	
10a	LA-O^{t}Bu	176.4 ± 16	
10a'	LA-OH	-	-
11a'	LI-OH	623.2	
1a	LAI- O^{t}Bu	111.1 ± 18	19.4 ± 4.1 (at 100 μM)
1a'	LAI-OH	3.7 ± 0.4	0.38 ± 0.07 (at 5 μM)
1b	FLA- O^{t}Bu	54.9 ± 6	21.0 ± 2.1 (at 50 μM)
1b'	FLA-OH	46.5 ± 8	0.93 ± 0.3 (at 50 μM)
12	LA-LA	84.4 ± 10	

7.6 Conclusion

PPIs are important but under-exploited therapeutic targets. Many methods have been developed to identify small molecules that disrupt PPIs. Existing computational methods tend to focus on how small pharmacophores bind protein surfaces through analyses of protein surfaces and fragment methods. However, these methods are limited by the difficulties in clustering fragments into one molecule, and they are only applicable to certain protein targets. Thus, most active compounds have been found by screening huge libraries via high throughput screen (HTS). However, there are several problems for HTS; it is time-consuming, expensive, slow, unreliable, and the rates for finding active compounds are low. Therefore, more effective methods to design small molecules that mimic or disrupt PPIs are highly desirable.

Exploring **K**ey **O**rientations (EKO) is an algorithm that uses a virtual affinity approach to find PPIs for preferred conformations of semi-rigid molecular scaffolds that display amino acid side-chains. EKO works by identifying conformations that adopt the same orientation as one protein component at the interface via matching Cα and Cβ coordinates for three side-chains. This strategy is distinct from existing methods because it; (i) focuses on conformations of one molecule, not fragments from several molecules, (ii) mines the conformations for whole protein-protein interfaces, not for specific structures, active sites, or hot-spots, (iii) finds the best matching between three Cα – Cβ bond vectors of small molecule conformations and protein interface regions.

We applied EKO for pyrrolinone-pyrrolidine oligomer scaffold **1**. EKO found 58 different possible targets including heat-labile enterotoxin/cholera toxin (cholera), AICAR (cancer), GAPDH (neurodegenerative disease), HIV-1 protease (AIDS), and NOS (AIDS, neurodegenerative disease, and cancer). HIV-1 protease dimerization inhibition assays were carried out to seek the first evidence for experimental validation.

For the assays, compounds **1a** and **1a'** that were obtained by EKO, compounds **1b**, and **1b'** that were obtained by eKO, and their intermediates **9** – **12** were synthesized, and tested for FRET based assays. Three side-chains compounds **1a**, **1a'**, **1b**, and **1b'** showed better inhibition effects than two side-chains compounds **9** - **12**. Unprotected

forms **1a'** and **1b'** gave better activities than *C*-protected form **1a** and **1b**. Moreover, Zhang-Poorman analyses for the three side-chains compounds **1a**, **1a'**, **1b**, and **1b'** proved that the compounds are inhibitors for HIV-1 protease dimerization, not for the active site.

We developed the new method, EKO which can identify a small molecule being able to disrupt PPIs, and the biology data validated the function. We have already applied EKO to other scaffolds, and have found many interesting targets.

CHAPTER VIII

CONCLUSIONS

Originally, our group designed small molecules that mimic β-turn regions that are known to be hot-spots for binding of neurotrophins to Trk receptors. Our original leads, before I joined the group, were peptide-containing macrocycles, and I followed others in the group by relating these to minimalist mimics. Specifically, I made small molecules based on rigid scaffolds with triazole or pyrrole units that substitute for peptide backbones; these are only peptidomimetics because they displayed amino acid side-chains, i.e. they are minimalist mimics. Several triazole-based minimalist mimics were synthesized, then assembled into bivalent mimics via a combinatorial method.

Four bivalent peptidomimetics that were prepared in Chapter I selectively bound to TrkC receptor, and showed partial antagonistic effects for the TrkC receptor. It is important to find active compounds that selectively bind to Trk receptors, because the selective binding for each Trk receptor leads to different effects on cells such as growth and apoptosis. Moreover, such compounds can be very useful pharmacological probes that target neurodegenerative diseases or cancers. In fact, we applied another triazole-based bivalent mimics that selectively bind to TrkC receptor to targeted drug delivery on TrkC overexpressed cells as described in Chapter III. TrkC receptors are overexpressed in cancers such as prostate, medulloblastoma, neuroblastoma, malignant melanoma, and pancreatic cancer. Antiproliferative assays showed the compounds conjugated with cytotoxic drugs targeted TrkC cells selectively over wild-type cells; more cytotoxicity for TrkC cells was shown than for wild-type cells.

In previous studies, we had a target, Trk receptors, for our small molecules, and the compounds were designed to mimic specific secondary structure (in this case, β-turn) on neurotrophins that are natural ligands for Trk receptors. However, application of triazole-based bivalent peptidomimetics in Chapter IV was different. Although originally the triazole-based bivalent mimics were designed as mimics of β-turns, the

compounds were screened for various cancer cells via high-throughput screening (specifically through the Luciferase assay), and KB1005 in the library was found to target pancreatic cancer cells. Cancer cells express a complicated set of cell surface receptors. Therefore, we do not know through which pathway the compound targets the cancer cells. This result is significant because we were able to find a mimic for an unknown target without any structural information.

At this stage we became aware that some minimalist mimics could resemble more than one secondary structure, and this led us to develop the idea that some compounds could be "universal peptidomimetics". Universal peptidomimetics are small sets of such scaffolds designed to analog local pairs of amino acids including non-contiguous ones in any secondary structures. These types of compounds are especially useful for targets where binding conformations are unknown, because the compounds can mimic any structures in protein-protein interfaces. Therefore, the peptidomimetics libraries are useful for high-throughput screening against various targets; our previous study that found KB1005 as a target for pancreatic cancer cells via HTS proved this. Moreover, we performed a molecular modeling, density functional theory (DFT) calculations, and quantum molecular dynamics (QMD) for other new templates such as oxazoline-, pyrrole-, dyine- "kinked" and "linear" bistrizole-based peptidomimetics to validate our idea more.

In that work, we used Cβ - Cβ separations to give a "rough cut" to fit mimics to secondary structures. However, Cα - Cβ bond vectors are a more sophisticated standard because they depict how the side-chains project. Computational methods were required to match Cα - Cβ bond vectors for mimics and ideal secondary structures. We developed the Cα - Cβ vector-matching program to do this, and applied the program to new templates such as omegatides and pyrrolinone-pyrrolidine oligomers with three amino acid side-chains. We were especially interested in three side-chains because scaffolds with three side-chains roughly correspond to tripeptides; strands of three amino acids have been recognized minimal motifs for effective molecular interactions. Both

omegatides and pyrrolinone-pyrrolidine oligomers matched better with strand structures than for helices.

At this stage we became more convinced that structural information on protein-protein interactions (PPIs) had to be more useful in discovery of small "drug-like" pharmacological probes and pharmaceutical leads for these targets than it currently was. Consequently, we developed the Exploring Key Orientations (EKO) algorithm that uses a virtual affinity approach to find PPIs for preferred conformations of semi-rigid molecular scaffolds that display amino acid side-chains. This strategy is distinct from existing methods to find small molecules that can disturb PPIs because it; (i) focuses on conformations of one molecule, not fragments from several molecules, (ii) mines the conformations for whole protein-protein interfaces, not for specific structures, active sites, or hot-spots, (iii) finds the best matching between three Cα – Cβ bond vectors of small molecule conformations and protein interface regions. We applied EKO to pyrrolinone-pyrrolidine oligomers to find targets, and she found several interesting targets relative to diseases, such as heat-labile enterotoxin/cholera toxin (cholera), AICAR (cancer), GAPDH (neurodegenerative disease), HIV-1 protease (AIDS), and NOS (AIDS, neurodegenerative disease, and cancer). HIV-1 dimerization inhibition assays and Zhang-Poorman kinetic assays were performed to validate our hypothesis for one of these targets (HIV-1 protease), and the results showed that pyrrolinone-pyrrolidine oligomer derivatives inhibited HIV-1 dimerization.

Others in the group are already beginning to design and synthesize compounds that conform to our chemotype guidelines for small molecules to perturb protein protein interactions. It is becoming clear that the pyrrolinone-pyrrolidine oligomer is not an optimal design mainly because we cannot make the compounds efficiently enough to explore all the targets that EKO has exposed. Consequently, I expect better scaffold designs and more topical PPI targets than HIV-1 protease to emerge from the future work. My studies are simply the first of what we anticipate will be a much longer-term program on data mining small molecules designed for PPIs.

REFERENCES

(1) Ripka, A. S.; Rich, D. H. *Curr. Opin. Chem. Biol.* **1998**, *2*, 441-452.

(2) Ahn, J.-M.; Boyle, N. A.; MacDonald, M. T.; Janda, K. D. *Mini Rev. Med. Chem.* **2002**, *2*, 463-473.

(3) Farmer, P. S. In *Drug Design*; Ariens, E. J., Ed.; Academic Press: San Diego, 1980; Vol. 10, p 119-143.

(4) Kale, S. S.; Chavan, S. T.; Sabharwal, S. G.; Puranik, V. G.; Sanjayan, G. J. *Org. Biomol. Chem.* **2011**, *9*, 7300-7302.

(5) Patgiri, A.; Jochim, A. L.; Arora, P. S. *Acc. Chem. Res.* **2008**, *41*, 1289-1300.

(6) Reddy, D. S.; Velde, D. V.; Aube, J. *J. Org. Chem.* **2004**, *69*, 1716-1719.

(7) Conte, L. L.; Chothia, C.; Janin, J. *J. Mol. Biol.* **1999**, *285*, 2177-2198.

(8) Ernst, J. T.; Becerril, J.; Park, H.; Yin, H.; Hamilton, A. D. *Angew. Chem. Int. Ed.* **2003**, *42*, 535-539.

(9) Davis, J. M.; Truong, A.; Hamilton, A. D. *Org. Lett.* **2005**, *7*, 5405-5408.

(10) Kim, I. C.; Hamilton, A. D. *Org. Lett.* **2006**, *8*, 1751-1754.

(11) Davis, J. M.; Tsou, L. K.; Hamilton, A. D. *Chem. Soc. Rev.* **2007**, *36*, 326-334.

(12) Ahn, J.-M.; Han, S.-Y. *Tetrahedron Letters* **2007**, *48*, 3543-3547.

(13) Marimganti, S.; Cheemala, M. N.; Ahn, J.-M. *Org. Lett.* **2009**, *11*, 4418-4421.

(14) Hirschmann, R.; Nicolaou, K. C.; Pietranico, S.; Salvino, J.; Leahy, E. M.; Sprengeler, P. A.; Furst, G.; Smith III, A. B. *J. Am. Chem. Soc.* **1992**, *114*, 9217-9218.

(15) Hirschmann, R.; Nicolaou, K. C.; Pietranico, S.; Leahy, E. M.; Salvino, J.; Arison, B.; Cichy, M. A.; Spoors, P. G.; Shakespeare, W. C.; Sprengeler, P. A.; Hamley, P.; Smith III, A. B.; Reisine, T.; Raynor, K.; Maechler, L.; Donaldson, C.;

Vale, W.; Freidinger, R. M.; Cascieri, M. R.; Strader, C. D. *J. Am. Chem. Soc.* **1993**, *115*, 12550-12568.

(16) Hirschmann, R.; Sprengeler, P. A.; Kawasaki, T.; Leahy, J. W.; Shakespeare, W. C.; Amos B. Smith, I. *J. Am. Chem. Soc.* **1992**, *114*, 9699-9701.

(17) Mowery, B. P.; Prasad, V.; Kenesky, C. S.; Angeles, A. R.; Taylor, L. L.; Feng, J.-J.; Chen, W.-L.; Lin, A.; Cheng, F.-C.; Smith, A. B., III; Hirschmann, R. *Org. Lett.* **2006**, *8*, 4397-4400.

(18) Angelo, N. G.; Arora, P. S. *J. Am. Chem Soc.* **2005**, *127*, 17134-17135.

(19) Meireles, L. M. C.; Mustata, G. *Curr. Top. Med. Chem.* **2011**, *11*, 248-257.

(20) Macarron, R.; Banks, M. N.; Bojanic, D.; Burns, D. J.; Cirovic, D. A.; Garyantes, T.; Green, D. V. S.; Hertzberg, R. P.; Janzen, W. P.; Paslay, J. W.; Schopfer, U.; Sittampalam, G. S. *Nat. Rev. Drug Discovery* **2011**, *10*, 188-195.

(21) Erlanson, D. A. *Curr. Opin. Biotechnol.* **2006**, *17*, 643-652.

(22) Erlanson, D. A.; Wells, J. A.; Braisted, A. C. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 199-223.

(23) Pickett, S. D.; Sherborne, B. S.; Wilkinson, T.; Bennett, J.; Borkakoti, N.; Broadhurst, M.; Hurst, D.; Kilford, I.; McKinnell, M.; Jones, P. S. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1691-1694.

(24) Beevers, R. E.; Buckley, G. M.; Davies, N.; Fraser, J. L.; Galvin, F. C.; Hannah, D. R.; Haughan, A. F.; Jenkins, K.; Mack, S. R.; Pitt, W. R.; Ratcliffe, A. J.; Richard, M. D.; Sabin, V.; Sharpe, A.; Williams, S. C. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 2535-2538.

(25) Beevers, R. E.; Buckley, G. M.; Davies, N.; Fraser, J. L.; Galvin, F. C.; Hannah, D. R.; Haughan, A. F.; Jenkins, K.; Mack, S. R.; Pitt, W. R.; Ratcliffe, A. J.; Richard, M. D.; Sabin, V.; Sharpe, A.; Williams, S. C. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 2539-2542.

(26) Hajduk, P. J.; Greer, J. *Nat. Rev. Drug Discovery* **2007**, *6*, 211-219.

(27) Nienaber, V. L.; Richardson, P. L.; Klighofer, V.; Bouska, J. J.; Giranda, V. L.; Greer, J. *Nat. Biotechnol.* **2000**, *18*, 1105-1108.

(28) Sun, H. *Curr. Med. Chem.* **2008**, *15*, 1018-1024.

(29) Kroemer, R. T. *Curr. Protein Pept. Sci.* **2007**, *8*, 312-328.

(30) Brown, J. B.; Okuno, Y. *Chem. Biol.* **2012**, *19*, 23-28.

(31) Stahura, F. L.; Bajorath, J. *Curr. Pharm. Des.* **2005**, *11*, 1189-1202.

(32) Rognan, D. *Br. J. Pharmacol.* **2007**, *152*, 38-52.

(33) Yamanishi, Y.; Kotera, M.; Kanehisa, M.; Goto, S. *Bioinformatics* **2010**, *26*, i246-i254.

(34) Chao, M. V. *Nat. Rev. Neurosci.* **2003**, *4*, 299-309.

(35) Longo, F. M.; Xie, Y.; Massa, S. M. *Curr. Med. Chem. Cent. Nerv. Syst. Agents* **2005**, *5*, 29-41.

(36) Pattarawarapan, M.; Burgess, K. *J. Med. Chem.* **2003**, *46*, 5277-5291.

(37) Huang, E. J.; Reichardt, L. F. *Annu. Rev. Biochem.* **2003**, *72*, 609-642.

(38) Saragovi, H. U.; Burgess, K. *Exp. Opin. Ther. Patents* **1999**, *9*, 737-751.

(39) Rubin, J. B.; Segal, R. A. *Cancer Treat. Res.* **2003**, *115*, 1-18.

(40) Bassili, M.; Birman, E.; Schor, N. F.; Saragovi, H. U. *Cancer Chemother. Pharmacol.* **2010**, *65*, 1047-1056.

(41) Pollack, S. J.; Harper, S. J. *Curr. Drug Targets: CNS Neurol. Disord.* **2002**, *1*, 59-80.

(42) Jang, S.-W.; Okada, M.; Sayeed, I.; Xiao, G.; Stein, D.; Jin, P.; Ye, K. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 16329-16334.

(43) Tapley, P.; Lamballe, F.; Barbacid, M. *Oncogene* **1992**, *7*, 371-381.

(44) Feng, Y.; Pattarawarapan, M.; Wang, Z.; Burgess, K. *Org. Lett.* **1999**, *1*, 121-124.

(45) Pattarawarapan, M.; Zaccaro, M. C.; Saragovi, U.; Burgess, K. *J. Med. Chem.* **2002**, *45*, 4387-4390.

(46) Lee, H. B.; Zaccaro, M. C.; Pattarawarapan, M.; Roy, S.; Saragovi, H. U.; Burgess, K. *J. Org. Chem.* **2004**, *69*, 701713.

(47) Angell, Y.; Chen, D.; Brahimi, F.; Saragovi, H. U.; Burgess, K. *J. Am. Chem. Soc.* **2008**, *130*, 556-565.

(48) Chen, D.; Brahimi, F.; Angell, Y.; Li, Y.-C.; Moscowicz, J.; Saragovi, H. U.; Burgess, K. *ACS Chem. Biol.* **2009**, *4*, 769-781.

(49) Liu, J.; Brahimi, F.; Saragovi, H. U.; Burgess, K. *J. Med. Chem.* **2010**, *53*, 5044-5048.

(50) Maliartchouk, S.; Feng, Y.; Ivanisevic, L.; Debeir, T.; Cuello, A. C.; Burgess, K.; Saragovi, H. U. *Mol. Pharmacol.* **2000**, *57*, 385-391.

(51) Bruno, M. A.; Clarke, P. B. S.; Seltzer, A.; Quirion, R.; Burgess, K.; Cuello, A. C.; Saragovi, H. U. *J. Neurosci.* **2004**, *24*, 8009-8018.

(52) Ko, E.; Ling, J.; Perez, L. M.; Lu, G.; Schaefer, A.; Burgess, K. *J. Am. Chem. Soc.* **2011**, *133*, 462-477.

(53) Capella, L.; Montevecchi, P. C.; Nanni, D. *J. Org. Chem.* **1994**, *59*, 3368-3374.

(54) Dias, L. C.; de Oliveira, L. G.; De Sousa, M. A.; Ellensohn, R. M. *ARKIVOC* **2005**, 62-87.

(55) Chaume, G.; Kuligowski, C.; Bezzenine-Laffolee, S.; Ricard, L.; Pancrazi, A.; Ardisson, J. *Synthesis* **2004**, 3029-3036.

(56) Alper, P. B.; Hung, S.-C.; Wong, C.-H. *Tetrahedron Letters* **1996**, *37*, 6029-6032.

(57) Dumas, D. J. *J. Org. Chem.* **1988**, *53*, 4650-4653.

(58) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. *Angew. Chem. Int. Ed.* **2002**, *41*, 2596-2599.

(59) Maliartchouk, S.; Debeir, T.; Beglova, N.; Cuello, A.; Gehring, K.; Saragovi, H. *J. Biol. Chem.* **2000**, *275*, 9946-9956.

(60) Maliartchouk, S.; Saragovi, H. U. *J. Neurosci.* **1997**, *17*, 6031-6037.

(61) Brahimi, F.; Malakhov, A.; Lee, H. B.; Pattarawarapan, M.; Ivanisevic, L.; Burgess, K.; Saragovi, H. U. *Peptides* **2009**, *30*, 1833-1839.

(62) Zaccaro, M. C.; Lee, B. H.; Pattarawarapan, M.; Xia, Z.; Caron, A.; L' Heureux, P.-J.; Bengio, Y.; Burgess, K.; Saragovi, H. U. *Chem. Biol.* **2005**, *12*, 1015-1028.

(63) Lu, Y.; Yang, J.; Sega, E. *AAPS Journal* **2006**, *8*, E466-E478.

(64) Wu, H.-C.; Chang, D.-K.; Huang, C.-T. *J. Cancer Mol.* **2006**, *2*, 57-66.

(65) Skipper, H. E.; Montgomery, J. A.; Thomson, J. R.; Schabel, F. M., Jr. *Cancer Res.* **1959**, *19*, 287-308.

(66) Schabel, F. M., Jr.; Montgomery, J. A.; Skipper, H. E.; Laster, W. R., Jr.; Thomson, J. R. *Cancer Res.* **1961**, *21*, 690-699.

(67) Nielsen, O. H.; Vainer, B.; Rask-Madsen, J. *Aliment. Pharmacol. Ther.* **2001**, *15*, 1699-1708.

(68) Lim, S. H.; Wu, L.; Burgess, K.; Lee, H. B. *Anti-Cancer Drugs* **2009**, *20*, 461-468.

(69) Wu, L.; Burgess, K. *J. Org. Chem.* **2008**, *73*, 8711-8718.

(70) Carroll, F. I.; Philip, A. *J. Org. Chem.* **1968**, *33*, 3776-3779.

(71) Nelson, J. A.; Vidale, E. *Cancer Res.* **1986**, *46*, 137-140.

(72) Butowt, R.; von Bartheld, C. S. *J. Neurosci.* **2001**, *21*, 8915-8930.

(73) Shi, Q.; Nguyen, A. T.; Angell, Y.; Deng, D.; Na, C.-R.; Burgess, K.; Roberts, D. D.; Brunicardi, F. C.; Templeton, N. S. *Gene Ther.* **2010**, *17*, 1085-1097.

(74) Handl, H. L.; Gillies, R. J. *Life Sci.* **2005**, *77*, 361-371.

(75) De Silva, C. R.; Vagner, J.; Lynch, R.; Gillies, R. J.; Hruby, V. J. *Anal. Biochem.* **2010**, *398*, 15-23.

(76) Wu, F.-B.; Han, S.-Q.; Xu, T.; He, Y.-F. *Anal. Biochem.* **2003**, *314*, 87-96.

(77) Handl, H. L.; Vagner, J.; Yamamura, H. I.; Hruby, V. J.; Gillies, R. J. *Anal. Biochem.* **2004**, *330*, 242-350.

(78) Handl, H. L.; Vagner, J.; Yamamura, H. I.; Hruby, V. J.; Gillies, R. J. *Anal. Biochem.* **2005**, *343*, 299-307.

(79) Xu, L.; Vagner, J.; Alleti, R.; Rao, V.; Jagadish, B.; Morse, D. L.; Hruby, V. J.; Gillies, R. J.; Mash, E. A. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 2489-2492.

(80) Pattarawarapan, M.; Reyes, S.; Xia, Z.; Zaccaro, M. C.; Saragovi, H. U.; Burgess, K. *J. Med. Chem.* **2003**, *46*, 3565-3567.

(81) Liu, J.; Brahimi, F.; Saragovi, H. U.; Burgess, K. *J. Med. Chem.* **2010**, *53*, 5044-5048.

(82) Ko, E.; Liu, J.; Burgess, K. *Chem. Soc. Rev.* **2011**, *40*, 4411-4421.

(83) Accelrys, I. 2010.

(84) Frisch, M. J. T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A.; Gaussian, Inc.: Wallingford, CT, 2004, p Gaussian 03.

(85) Pettitt, B. M.; Matsunaga, T.; Al-Obeidi, F.; Gehrig, C.; Hruby, V. J.; Karplus, M. *Biophys. J.* **1991**, *60*, 1540-1544.

(86) O'Connor, S. D.; Smith, P. E.; Al-Obeidi, F.; Pettitt, B. M. *J. Med. Chem.* **1992**, *35*, 2870-2881.

(87) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781-1802.

(88) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33-8.

(89) Hetzheim, A.; Moeckel, K. *Adv. Heterocycl. Chem.* **1966**, *7*, 183-224.

(90) Hill, J. *Comprehensive Heterocyclic Chemistry II* **1996**, *4*, 267-287.

(91) Raman, K.; Parmar, S. S.; Salzman, S. K. *J. Pharm. Sci.* **1989**, *78*, 999-1002.

(92) Mazzone, G.; Bonina, F.; Puglisi, G.; Panico, A. M.; Arrigo Reina, R. *Farmaco, Edizione Scientifica* **1984**, *39*, 414-420.

(93) Shi Shun, A. L. K.; Tykwinski, R. R. *Angew. Chem. Int. Ed.* **2006**, *45*, 1034-1057.

(94) Yun, H.; Danishefsky, S. J. *J. Org. Chem.* **2003**, *68*, 4519-4522.

(95) Ohira, S. *Synth. Commun.* **1989**, *19*, 561-564.

(96) Roth, G. J.; Liepold, B.; Muller, S. G.; Bestmann, H. J. *Synthesis* **2004**, 59-62.

(97) Siemsen, P.; Livingston, R. C.; Diederich, F. *Angew. Chem. Int. Ed.* **2000**, *39*, 2632-2657.

(98) Baltazzi, E.; Krimen, L. I. *Chem. Rev.* **1963**, *63*, 511-556.

(99) Balme, G. *Angew. Chem. Int. Ed.* **2004**, *43*, 6238-6241.

(100) Ferreira, V. F.; De Souza, M. C. B. V.; Cunha, A. C.; Pereira, L. O. R.; Ferreira, M. L. G. *Org. Prep. Proced. Int.* **2001**, *33*, 411-454.

(101) Schmuck, C.; Rupprecht, D. *Synthesis* **2007**, *20*, 3095-3110.

(102) Chakraborty, T. K.; Krishna Mohan, B.; Kiran Kumar, S.; Kunwar, A. C. *Tetrahedron Letters* **2002**, *43*, 2589-2592.

(103) Alongi, M.; Minetto, G.; Taddei, M. *Tetrahedron Letters* **2005**, *46*, 7069.

(104) Minchev, I.; Vladimirova, S.; Vezenkov, L.; Bijev, A.; Moussis, V.; Nikolaeva-Glomb, L.; Tsikaris, V.; Czeuz, M.; Galabov, A. *Protein Pept. Lett.* **2007**, *14*, 917-922.

(105) Yavari, I.; Kowsari, E. *Synlett* **2008**, 897-899.

(106) Zhang, Z.; Zhang, J.; Tan, J.; Wang, Z. *J. Org. Chem.* **2008**, *73*, 5180-5182.

(107) Alizadeh, A.; Hosseinpour, R.; Rostamnia, S. *Synthesis* **2008**, 2462-2466.

(108) Gouault, N.; Le Roch, M.; Cornee, C.; David, M.; Uriac, P. *J. Org. Chem.* **2009**, *74*, 5614-5617.

(109) Mishriky, N.; Asaad, F. M.; Ibrahim, Y. A.; Girgis, A. S. *Die Pharmazie* **1998**, *53*, 607-611.

(110) Yamawaki, I.; Matsushita, Y.; Asaka, N.; Ohmori, K.; Nomura, N.; Ogawa, K. *Eur. J. Med. Chem.* **1993**, *28*, 481-498.

(111) Varasi, M.; Heidempergher, F.; Caccia, C.; Salvati, P. P*harmacia S.p.A., Italy* 1995, p 29.

(112) Balasubramaniam, S.; Aidhen, I. S. *Synthesis* **2008**, *2008*, 3707-3738.

(113) Palmer, J. T.; Rasnick, D.; Klaus, J. L.; Bromme, D. *J. Med. Chem.* **1995**, *38*, 3193-3196.

(114) Ghosh, A. K.; Xi, K.; Ratia, K.; Santarsiero, B. D.; Fu, W.; Harcourt, B. H.; Rota, P. A.; Baker, S. C.; Johnson, M. E.; Mesecar, A. D. *J. Med. Chem.* **2005**, *48*, 6767-6771.

(115) Smith, C. R.; Bunnelle, E. M.; Rhodes, A. J.; Sarpong, R. *Org. Lett.* **2007**, *9*, 1169-1171.

(116) Sakaitani, M.; Ohfune, Y. *J. Org. Chem.* **1990**, *55*, 870-876.

(117) Youle, R. J.; Strasser, A. *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 47-59.

(118) Lamark, T.; Perander, M.; Outzen, H.; Kristiansen, K.; Overvatn, A.; Michaelsen, E.; Bjorkoy, G.; Johansen, T. *J. Biol. Chem.* **2003**, *278*, 34568-34581.

(119) Ung, P.; Winkler, D. A. *J. Med. Chem.* **2011**, *54*, 1111-1125.

(120) Keskin, O.; Gursoy, A.; Ma, B.; Nussinov, R. *Chem. Rev.* **2008**, *108*, 1225-1244.

(121) Clackson, T.; Wells, J. A. *Science* **1995**, *267*, 383-386.

(122) Keskin, O.; Ma, B.; Nussinov, R. *J. Mol. Biol.* **2005**, *345*, 1281-1294.

(123) Guharoy, M.; Chakrabarti, P. *Bioinformatics* **2007**, *23*, 1909-1918.

(124) Reynolds, C. H.; Bembenek, S. D.; Tounge, B. A. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 4258-4261.

(125) Neduva, V.; Russell, R. B. *Nucleic Acids Res.* **2006**, *34*, W350-W355.

(126) Wallace, A. C.; Borkakoti, N.; Thornton, J. M. *Protein Science* **1997**, *6*, 2308-2323.

(127) Fedoseyenko, D.; Raghuraman, A.; Ko, E.; Burgess, K. *Org. Biomol. Chem.* **2012**, *10*, 921-924.

(128) Raghuraman, A.; Ko, E.; Perez, L. M.; Ioerger, T. R.; Burgess, K. *J. Am. Chem. Soc.* **2011**, *133*, 12350-12353.

(129) Wells, J. A.; McClendon, C. L. *Nature* **2007**, *450*, 1001-1009.

(130) Macarron, R. *Drug Discov Today* **2006**, *11*, 277-279.

(131) Groenen, L. C.; Nice, E. C.; Burgess, A. W. *Growth Factors* **1994**, *11*, 235-257.

(132) Ogiso, H.; Ishitani, R.; Nureki, O.; Fukai, S.; Yamanaka, M.; Kim, J.-H.; Saito, K.; Sakamoto, A.; Inoue, M.; Shirouzu, M.; Yokoyama, S. *Cell* **2002**, *110*, 775-787.

(133) Lauri, G.; Bartlett, P. A. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 51-66.

(134) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *J. Mol. Biol.* **1982**, *161*, 269-288.

(135) Levy, E. D.; Pereira-Leal, J. B.; Chothia, C.; Teichmann, S. A. *PLoS Comput. Biol.* **2006**, *2*, 1395-1406.

(136) Fan, E.; Merritt, E. A.; Zhang, Z.; Pickens, J. C.; Roach, C.; Ahn, M.; Hol, W. G. J. *Acta Crystallogr., Sect. D Biol. Crystallogr.* **2001**, *D57*, 201-212.

(137) Wolan, D. W.; Greasley, S. E.; Beardsley, G. P.; Wilson, I. A. *Biochemistry* **2002**, *41*, 15505-15513.

(138) Axelrod, H. L.; McMullan, D.; Krishna, S. S.; Miller, M. D.; Elsliger, M.-A.; Abdubek, P.; Ambing, E.; Astakhova, T.; Carlton, D.; Chiu, H.-J.; Clayton, T.; Duan, L.; Feuerhelm, J.; Grzechnik, S. K.; Hale, J.; Han, G. W.; Haugen, J.; Jaroszewski, L.; Jin, K. K.; Klock, H. E.; Knuth, M. W.; Koesema, E.; Morse, A. T.; Nigoghossian, E.; Okach, L.; Oommachen, S.; Paulsen, J.; Quijano, K.; Reyes, R.; Rife, C. L.; van den Bedem, H.; Weekes, D.; White, A.; Wolf, G.; Xu, Q.; Hodgson, K. O.; Wooley, J.; Deacon, A. M.; Godzik, A.; Lesley, S. A.; Wilson, I. A. *Proteins Struct., Funct., Bioinf.* **2008**, *71*, 1042-1049.

(139) Mercer, W. D.; Winn, S. I.; Watson, H. C. *J. Mol. Biol.* **1976**, *104*, 277-283.

(140) Chuang, D.-M.; Hough, C.; Senatorov, V. V. *Annu. Rev. Pharmacol. Toxicol.* **2005**, *45*, 269-290.

(141) Kohl, N. E.; Emini, E. A.; Schleif, W. A.; Davis, L. J.; Heimbach, J. C.; Dixon, R. A. F.; Scolnick, E. M.; Sigal, I. S. *Proc. Natl. Acad. Sci.* **1988**, *85*, 4686-4690.

(142) Crane, B. R.; Arvai, A. S.; Ghosh, D. K.; Wu, C.; Getzoff, E. D.; Stuehr, D. J.; Tainer, J. A. *Science* **1998**, *279*, 2121-2126.

(143) De Clercq, E. *Nat. Rev. Microbiol.* **2004**, *2*, 704-720.

(144) Carr, A. *Nat. Rev. Drug Discovery* **2003**, *2*, 624-634.

(145) Darke, P. L.; Hall, D. L.; Jordan, S. P.; Zugay, J. A.; Shafer, J. A.; Kuo, L. C. *Biochemistry* **1994**, *33*, 98-105.

(146) Zhang, Z. Y.; Poorman, R. A.; Maggiora, L. L.; Heinrikson, R. L.; Kezdy, F. J. *J. Biol. Chem.* **1991**, *266*, 15591-15594.

(147) Todd, M. J.; Semo, N.; Freire, E. *J. Mol. Biol.* **1998**, *283*, 475-488.

(148) Davis, D. A.; Dorsey, K.; Wingfield, P. T.; Stahl, S. J.; Kaufman, J.; Fales, H. M.; Levine, R. L. *Biochemistry* **1996**, *35*, 2482-2488.

(149) Schramm, H. J.; Nakashima, H.; Schramm, W.; Wakayama, H.; Yamamoto, N. *Biochem. Biophys. Res. Commun.* **1991**, *179*, 847-851.

(150) Schramm, H. J.; Billich, A.; Jaeger, E.; Rucknagel, K. P.; Arnold, G.; Schramm, W. *Biochem. Biophys. Res. Commun.* **1993**, *194*, 595-600.

(151) Franciskovich, J.; Houseman, K.; Mueller, R.; Chmielewski, J. *Bioorg. Med. Chem. Lett.* **1993**, *3*, 765-768.

(152) Zutshi, R.; Franciskovich, J.; Shultz, M.; Schweitzer, B.; Bishop, P.; Wilson, M.; Chmielewski, J. *J. Am. Chem. Soc.* **1997**, *119*, 4841-4845.

(153) Shultz, M. D.; Ham, Y.-W.; Lee, S.-G.; Davis, D. A.; Brown, C.; Chmielewski, J. *J. Am. Chem. Soc.* **2004**, *126*, 9886-9887.

(154) Hwang, Y. S.; Chmielewski, J. *J. Med. Chem.* **2005**, *48*, 2239-2242.

(155) Davis, D. A.; Tebbs, I. R.; Daniels, S. I.; Stahl, S. J.; Kaufman, J. D.; Wingfield, P.; Bowman, M. J.; Chmielewski, J.; Yarchoan, R. *Biochem. J.* **2009**, *419*, 497-506.

(156) Davis, D. A.; Brown, C. A.; Singer, K. E.; Wang, V.; Kaufman, J.; Stahl, S. J.; Wingfield, P.; Maeda, K.; Harada, S.; Yoshimura, K.; Kosalaraksa, P.; Mitsuya, H.; Yarchoan, R. *Antiviral Res.* **2006**, *72*, 89-99.

(157) Lee, S.-G.; Chmielewski, J. *ChemBioChem* **2010**, *11*, 1513-1516.

(158) Vidu, A.; Dufau, L.; Bannwarth, L.; Soulier, J.-L.; Sicsic, S.; Piarulli, U.; Reboud-Ravaux, M.; Ongeri, S. *ChemMedChem* **2010**, *5*, 1899-1906.

(159) Koh, Y.; Matsumi, S.; Das, D.; Amano, M.; Davis, D. A.; Li, J.; Leschenko, S.; Baldridge, A.; Shioda, T.; Yarchoan, R.; Ghosh, A. K.; Mitsuya, H. *J. Biol. Chem.* **2007**, *282*, 28709-28720.

(160) Kim, E. E.; Baker, C. T.; Dwyer, M. D.; Murcko, M. A.; Rao, B. G.; Tung, R. D.; Navia, M. A. *J. Am. Chem. Soc.* **1995**, *117*, 1181-1182.

(161) Reddy, G. S. K. K.; Ali, A.; Nalam, M. N. L.; Anjum, S. G.; Cao, H.; Nathans, R. S.; Schiffer, C. A.; Rana, T. M. *J. Med. Chem.* **2007**, *50*, 4316-4328.

(162) Ghosh, A. K.; Xi, K.; Ratia, K.; Santarsiero, B. D.; Fu, W.; Harcourt, B. H.; Rota, P. A.; Baker, S. C.; Johnson, M. E.; Mesecar, A. D. *J. Med. Chem.* **2005**, *48*, 6767-6771.

(163) Tello-Aburto, R.; Olivo, H. F. *Org. Lett.* **2008**, *10*, 2191-2194.

(164) Lencer, W. I.; Hirst, T. R.; Holmes, R. K. *Biochim. Biophys. Acta, Mol. Cell Res.* **1999**, *1450*, 177-190.

(165) Sixma, T. K.; Pronk, S. E.; Kalk, K. H.; Wartna, E. S.; Van Zanten, B. A. M.; Witholt, B.; Hol, W. G. J. *Nature* **1991**, *351*, 371-377.

(166) Fan, E.; Merritt, E. A.; Zhang, Z.; Pickens, J. C.; Roach, C.; Ahn, M.; Hol, W. G. J. *Acta Crystallogr., Sect. D Biol. Crystallogr.* **2001**, *D57*, 201-212.

(167) Aman, A. T.; Fraser, S.; Merritt, E. A.; Rodigherio, C.; Kenny, M.; Ahn, M.; Hol, W. G. J.; Williams, N. A.; Lencer, W. I.; Hirst, T. R. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 8536-8541.

(168) Lesieur, C.; Cliff, M. J.; Carter, R.; James, R. F. L.; Clarke, A. R.; Hirst, T. R. *J. Biol. Chem.* **2002**, *277*, 16697-16704.

(169) Greasley, S. E.; Horton, P.; Ramcharan, J.; Beardsley, G. P.; Benkovic, S. J.; Wilson, I. A. *Nat. Struct. Biol.* **2001**, *8*, 402-406.

(170) Axelrod, H. L.; McMullan, D.; Krishna, S. S.; Miller, M. D.; Elsliger, M.-A.; Abdubek, P.; Ambing, E.; Astakhova, T.; Carlton, D.; Chiu, H.-J.; Clayton, T.; Duan, L.; Feuerhelm, J.; Grzechnik, S. K.; Hale, J.; Han, G. W.; Haugen, J.; Jaroszewski, L.; Jin, K. K.; Klock, H. E.; Knuth, M. W.; Koesema, E.; Morse, A. T.; Nigoghossian, E.; Okach, L.; Oommachen, S.; Paulsen, J.; Quijano, K.; Reyes, R.; Rife, C. L.; van den Bedem, H.; Weekes, D.; White, A.; Wolf, G.; Xu, Q.; Hodgson, K. O.; Wooley, J.; Deacon, A. M.; Godzik, A.; Lesley, S. A.; Wilson, I. A. *Proteins Struct., Funct., Bioinf.* **2008**, *71*, 1042-1049.

(171) Wolan, D. W.; Greasley, S. E.; Beardsley, G. P.; Wilson, I. A. *Biochemistry* **2002**, *41*, 15505-15513.

(172) Shim, J. H.; Wall, M.; Benkovic, S. J.; Diaz, N.; Suarez, D.; Merz, K. M., Jr. *J. Am. Chem. Soc.* **2001**, *123*, 4687-4696.

(173) Jackson, R. C.; Harkrader, R. J. In *Nucleosides and Cancer Treatment*; Tattersall, M. H. N., Fox, R. M., Eds.; Academic Press: Sydney, 1981, p 18-31.

(174) Berman, E. M.; Werbel, L. M. *J. Med. Chem.* **1991**, *34*, 479-485.

(175) Moras, R. G. *Cancer Treatment and Res.* **1991**, *58*, 65-69.

(176) Divekar, A. Y.; Hakala, M. T. *Mol. Pharmacol.* **1975**, *11*, 319-325.

(177) Erba, E.; Sen, S.; Sessa, C.; Vikhanskaya, F. L.; D'Incalci, M. *Br. J. Cancer* **1994**, *69*, 205-211.

(178) Beardsley, G. P.; Moroson, B. A.; Taylor, E. C.; Moran, R. G. *J. Biol. Chem.* **1989**, *264*, 328-333.

(179) Mendelsohn, L. G.; Shih, C.; Chen, V. J.; Habeck, L. L.; Gates, S. B.; Shackelford, K. A. *Semin. Oncol.* **1999**, *26*, 42-47.

(180) Tavassoli, A.; Benkovic, S. J. *Angew. Chem., Int. Ed.* **2005**, *44*, 2760-2763.

(181) Capps, K. J.; Humiston, J.; Dominique, R.; Hwang, I.; Boger, D. L. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2840-2844.

(182) Sirover, M. A. *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **1999**, *1432*, 159-184.

(183) Kragten, E.; Lalande, I.; Zimmermann, K.; Roggo, S.; Schindler, P.; Muller, D.; Van Oostrum, J.; Waldmeier, P.; Furstt, P. *J. Biol. Chem.* **1998**, *273*, 5821-5828.

(184) Cowan-Jacob, S. W.; Kaufmann, M.; Anselmo, A. N.; Stark, W.; Gruetter, M. G. *Acta Crystallogr., Sect. D Biol. Crystallogr.* **2003**, *D59*, 2218-2227.

(185) Berry, M. D. *J Psychiatry Neurosci* **2004**, *29*, 337-345.

(186) Berry, M. D.; Boulton, A. A. *J. Neurosci. Res.* **2000**, *60*, 150-154.

(187) Chuang, D.-M.; Hough, C.; Senatorov, V. V. *Annu. Rev. Pharmacol. Toxicol.* **2005**, *45*, 269-290.

(188) Meyer-Siegler, K.; Mauro, D. J.; Seal, G.; Wurzer, J.; DeRiel, J. K.; Sirover, M. A. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88*, 8460-8464.

(189) Hara, M. R.; Agrawal, N.; Kim, S. F.; Cascio, M. B.; Fujimuro, M.; Ozeki, Y.; Takahashi, M.; Cheah, J. H.; Tankou, S. K.; Hester, L. D.; Ferris, C. D.; Hayward, S. D.; Snyder, S. H.; Sawa, A. *Nat. Cell Biol.* **2005**, *7*, 665-674.

(190) Jenkins, J. L.; Tanner, J. J. *Acta Crystallogr., Sect. D Biol. Crystallogr.* **2006**, *D62*, 290-301.

(191) Schulze, H.; Schuler, A.; Stueber, D.; Doebeli, H.; Langen, H.; Huber, G. *J. Neurochem.* **1993**, *60*, 1915-1922.

(192) Carlile, G. W.; Chalmers-Redman, R. M. E.; Tatton, N. A.; Pong, A.; Borden, K. E.; Tatton, W. G. *Mol. Pharmacol.* **2000**, *57*, 2-12.

(193) Minton, A. P.; Wilf, J. *Biochemistry* **1981**, *20*, 4821-4826.

(194) Mercer, W. D.; Winn, S. I.; Watson, H. C. *J. Mol. Biol.* **1976**, *104*, 277-283.

(195) Dietmann, S.; Park, J.; Notredame, C.; Heger, A.; Lappe, M.; Holm, L. *Nucleic Acids Res.* **2001**, *29*, 55-57.

(196) Fuentes-Prior, P.; Salvesen, G. S. *Biochem. J.* **2004**, *384*, 201-232.

(197) Linton, S. D. *Curr. Top. Med. Chem.* **2005**, *5*, 1697-1716.

(198) Solary, E.; Eymin, B.; Droin, N.; Haugg, M. *Cell Biol. Toxicol.* **1998**, *14*, 121-132.

(199) Friedlander, R. M.; Gagliardini, V.; Rotello, R. J.; Yuan, J. *J. Exp. Med.* **1996**, *184*, 717-724.

(200) Friedlander, R. M. *Arch Neurol* **2000**, *57*, 1273-1276.

(201) Yuan, J.; Yankner, B. A. *Nature* **2000**, *407*, 802-809.

(202) Troy, C. M.; Salvesen, G. S. *J. Neurosci. Res.* **2002**, *69*, 145-150.

(203) Goldberg, Y. P.; Nicholson, D. W.; Rasper, D. M.; Kalchman, M. A.; Koide, H. B.; Graham, R. K.; Bromm, M.; Kazemi-Esfarjani, P.; Thornberry, N. A. *Nat. Genet.* **1996**, *13*, 442-449.

(204) Ona, V. O.; Li, M.; Vonsattel, J. P. G.; Andrews, L. J.; Khan, S. Q.; Chung, W. M.; Frey, A. S.; Menon, A. S.; Li, X.-J.; Stieg, P. E.; Yuan, J.; Penney, J. B.; Young, A. B.; Cha, J.-H. J.; Friedlander, R. M. *Nature* **1999**, *399*, 263-267.

(205) O'Brien, T.; Lee, D. *Mini-Rev. Med. Chem.* **2004**, *4*, 153-165.

(206) Boxer, M. B.; Shen, M.; Auld, D. S.; Wells, J. A.; Thomas, C. T. *A Small Molecule Inhibitor of Caspase 1*, **2011**, 1-19.

(207) O'Brien, T.; Linton, S. D. *Design of Caspase Inhibitors as Potential Clincal Agents*; CRC Press: Boca Raton, FL, 2008, p1-312.

(208) Nicholson, D. W. *Cell Death Differ.* **1999**, *6*, 1028-1042.

(209) Dobo, J.; Swanson, R.; Salvesen, G. S.; Olson, S. T.; Gettins, P. G. W. *J. Biol. Chem.* **2006**, *281*, 38781-38790.

(210) O'Brien, T.; Fahr, B. T.; Sopko, M. M.; Lam, J. W.; Waal, N. D.; Raimundo, B. C.; Purkey, H. E.; Pham, P.; Romanowski, M. J. *Acta Crystallogr., Sect. F Struct. Biol. Cryst. Commun.* **2005**, *F61*, 451-458.

(211) Scheer, J. M.; Romanowski, M. J.; Wells, J. A. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 7595-7600.

(212) Fu, G.; Chumanevich, A. A.; Agniswamy, J.; Fang, B.; Harrison, R. W.; Weber, I. T. *Apoptosis* **2008**, *13*, 1291-1302.

(213) Lomas, D. A.; Carrell, R. W. *Nat. Rev. Genet.* **2002**, *3*, 759-768.

(214) Knaupp Anja, S.; Bottomley Stephen, P. *IUBMB Life* **2009**, *61*, 1-5.

(215) Yamasaki, M.; Li, W.; Johnson, D. J. D.; Huntington, J. A. *Nature* **2008**, *455*, 1255-1258.

(216) Devlin, G. L.; Chow, M. K. M.; Howlett, G. J.; Bottomley, S. P. *J. Mol. Biol.* **2002**, *324*, 859-870.

(217) Zhou, A.; Carrell, R. W. *J. Mol. Biol.* **2008**, *375*, 36.

(218) Dafforn, T. R.; Mahadeva, R.; Elliott, P. R.; Sivasothy, P.; Lomas, D. A. *J. Biol. Chem.* **1999**, *274*, 9548.

(219) Chang, Y.-P.; Mahadeva, R.; Chang, W.-S. W.; Shukla, A.; Dafforn, T. R.; Chu, Y.-H. *Am. J. Respir. Cell Mol. Biol.* **2006**, *35*, 540-548.

(220) Mahadeva, R.; Dafforn, T. R.; Carrell, R. W.; Lomas, D. A. *J. Biol. Chem.* **2002**, *277*, 6771-6774.

(221) Skinner, R.; Chang, W.-S. W.; Jin, L.; Pei, X.; Huntington, J. A.; Abrahams, J.-P.; Carrell, R. W.; Lomas, D. A. *J. Mol. Biol.* **1998**, *283*, 9-14.

(222) Soto, C.; Sigurdsson, E. M.; Morelli, L.; Kumar, R. A.; Castano, E. M.; Frangione, B. *Nat. Med.* **1998**, *4*, 822-826.

APPENDIX A

GENERAL EXPERIMENTAL PROCEDURES

A. General Methods

All reactions were carried out under an atmosphere of dry nitrogen. Glassware was oven-dried prior to use. Unless otherwise indicted, common reagents or materials were obtained from commercial source and used without further purification. All α-amino acids used were of the L-configuration, except where otherwise indicated. Triethylamine (TEA) was obtained anhydrous by distillation over calcium hydride and tetrahydrofuran (THF) was distilled over sodium metal and benzophenone. Acetonitrile, dichloromethane, methanol and diethyl ether were dried by a Mbraun solvent drying system.

Flash column chromatography was performed using silica gel 60 (230-400 mesh). Analytical thin layer chromatography (TLC) was carried out on Merck silica gel plates with QF-254 indicator and visualized by UV. Optical rotations were measured on Jasco DIP-360 digital polarimeter. ^{1}H and ^{13}C NMR spectra were recorded on a Varian 300 (300 MHz ^{1}H; 75 MHz ^{13}C) or Varian 500 (500 MHz ^{1}H; 125 MHz ^{13}C) spectrometer at room temperature. Chemical shifts were reported in ppm relative to the residual $CDCl_3$ (δ 7.27 ppm ^{1}H; δ 77.0 ppm ^{13}C), CD_3OD (δ 3.31 ppm ^{1}H; δ 49.86 ppm ^{13}C), or d^6-DMSO (δ 2.49 ppm ^{1}H; δ 39.5 ppm ^{13}C). NMR chemical shifts were expressed in ppm relative to internal solvent peaks, and coupling constants were measured in Hz. (br = broad).

Analytical HPLC analyses were carried out on 150 x 4.6 mm C-18 column using gradient conditions (10 – 90% B, flow rate = 0.75 mL/min). Preparative HPLC was carried out on 100 x 21.2 mm C-18 column using gradient conditions (10 – 70% B, flow rate = 10.0 mL/min). The eluents used were: solvent A (H_2O with 0.1% TFA) and solvent B (CH_3CN with 0.1% TFA)

APPENDIX B

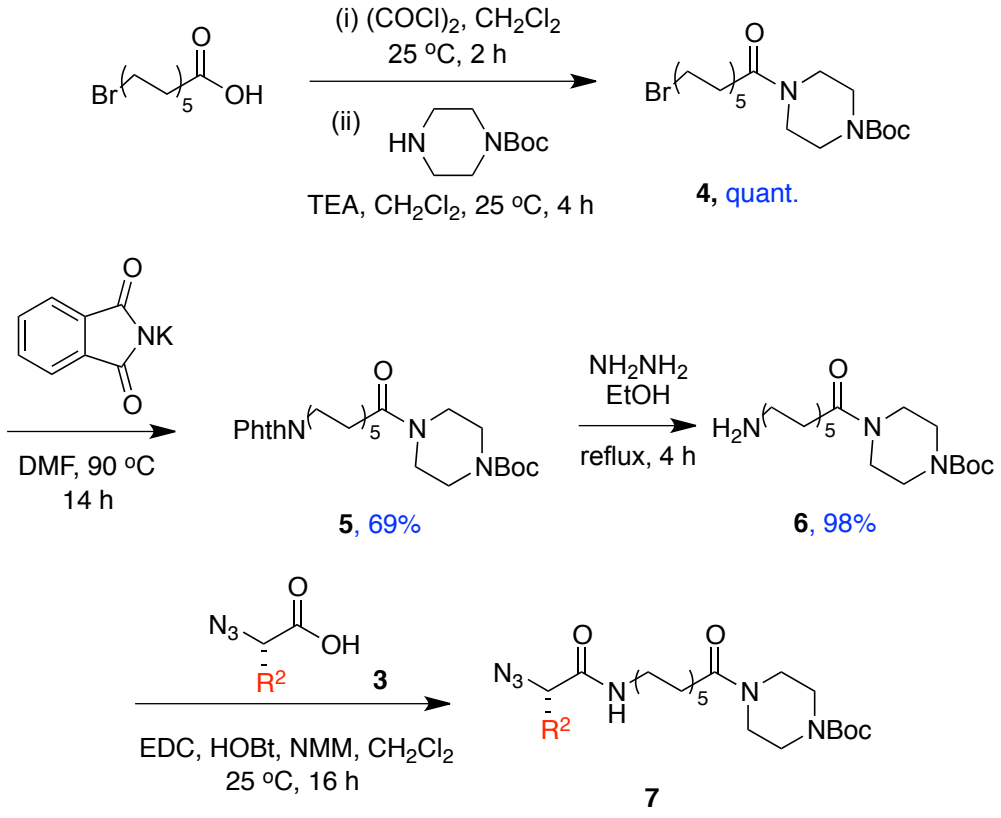
EXPERIMENTAL FOR CHAPTER II

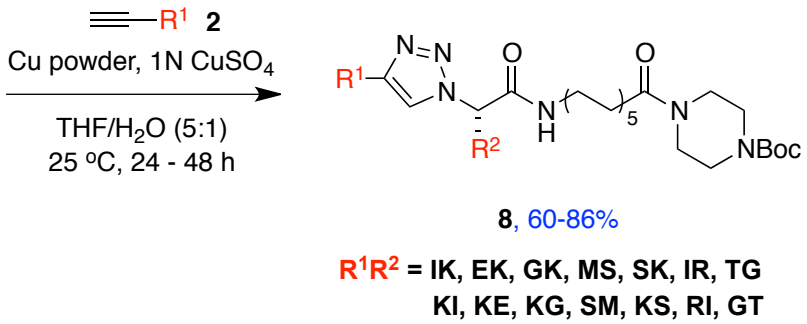
A. General Procedure and Preparation of Monovalent Mimics 8

Total 14 monovalent mimics were prepared by the following procedure in different sequence order of β-turn regions, e.g. *i*+1 and *i*+2 (TG) and *i*+2 and *i*+1 (GT). To the solution of 11-bromoundecanoic acid (1.0 equiv) in CH_2Cl_2 (1.0 M) was added oxalyl chloride (10.0 equiv). The mixture was stirred at 25 °C for 2 h, and then the solvent was removed. To remove the excess oxalyl chloride, CH_2Cl_2 was added in the resulting residue, and then was removed under vacuum (x2). The resulting residue was dissolved in CH_2Cl_2 (1.0 M), and then Boc-piperazine (1.0 equiv), and TEA (2.5 equiv) were added in the solution. The mixture was stirred at 25 °C for 4 h. After the solvent was removed under vacuum, the reaction mixture was diluted with H_2O and extracted with EtOAc, and washed with brine. The combined organic phases were dried over $MgSO_4$. After completely removing the solvent, the compound **4** was purified by flash chromatography (1:3 EtOAc/Hexanes). To the solution of compound **4** (1.0 equiv) in DMF (0.09 M) was added K_2CO_3 (1.05 equiv), and then the mixture was stirred at 90 °C for 14 h. The reaction mixture was diluted with H_2O and extracted with EtOAc. The combined organic phases were dried over Na_2SO_4. After completely removing the solvent, the compound **5** was purified by flash chromatography (2:5 EtOAc/Hexanes). The compound **6** (1.0 equiv) was dissolved in EtOH (0.16 M), and then hydrazine (4.0 equiv) was added in the solution. The mixture was refluxed at 90 °C for 4 h. After the reaction, the mixture was filtered, and then the solution was concentrated. Compound **6** was used for the next step without further purification. HOBt (1.1 equiv) and EDCI (1.1 equiv) were added to the solution of compound **3** (1.0 equiv) in CH_2Cl_2 (0.2 M) at 0 °C. The mixture was stirred at 0 °C for 15 min, and then NMM (2.0 equiv) and compound **6** (1.0 equiv) were added into the mixture. After the solution was warmed up to 25 °C, the mixture was stirred at the temperature for 16 h. The solvent was removed, and then the

resulting residue was diluted with H_2O and extracted with EtOAc. The combined organic layer was washed with 5% HCl (aq.), followed by 5% Na_2CO_3 (aq.), and brine, and then dried over Na_2SO_4. After completely removing the solvent, the compound **7** derivatives were purified by flash chromatography (EtOAc/Hexanes). Compound **7** (1.05 equiv) and **6** (1.0 equiv) were dissolved in THF:H_2O (5:1, 0.25 M), and then copper powder (1.0 equiv) and 1*N* $CuSO_4$ (aq., 0.01 equiv) were added. The mixture was stirred at 25 °C for 24 - 48 h. After the reaction, copper powder was filtered on celite with CH_2Cl_2 or EtOAc. The filtrate was washed with *sat.* NH_4Cl (aq.): NH_4OH (v:v=9:1), and brine. The organic layer was dried over Na_2SO_4. After completely removing the solvent, the compound **8** derivatives were purified by flash chromatography (MeOH:CH_2Cl_2).

Scheme S1. Preparation of Monovalent Mimics with Long Linkers.





8, 60-86%

R^1R^2 = IK, EK, GK, MS, SK, IR, TG
KI, KE, KG, SM, KS, RI, GT

B. General procedure and Preparation of Bivalent Mimics 1

General syntheses for bivalent mimics were modified from the published papers. Compounds **8** (1.0 equiv) were treated with 50% TFA in CH_2Cl_2 for 12h at 25 ºC and then the solvent was removed. The resulting residue was dissolved in THF (0.04 M), and DTAF (1.0 equiv or tag with biotin) and K_2CO_3 (4.0 equiv) were added. The suspension was stirred for 18 h at 25 ºC and then the solvent was removed. It was used for the next step without further purification. The resulting crude product was dissolved in DMSO (0.04 M), and then another deprotected compound and K_2CO_3 (4.0 equiv) were added. The suspension was stirred for 10 – 14 days at 25 °C. After monitoring staring materials' disappearance by analytical HPLC, the mixture was lyophilized to remove DMSO. The materials were re-dissolved in 1:1 mixture of H_2O/CH_3CN, and then purified by preparative HPLC to yield the final products. HCl (30 ml) and brine (30 ml), dried with Na_2SO_4, and concentrated under reduced pressure. Flash chromatography of the residue provided pure bivalent mimics **1**.

Scheme S2. Preparation of the Tagged Bivalent Peptidomimetics 1.

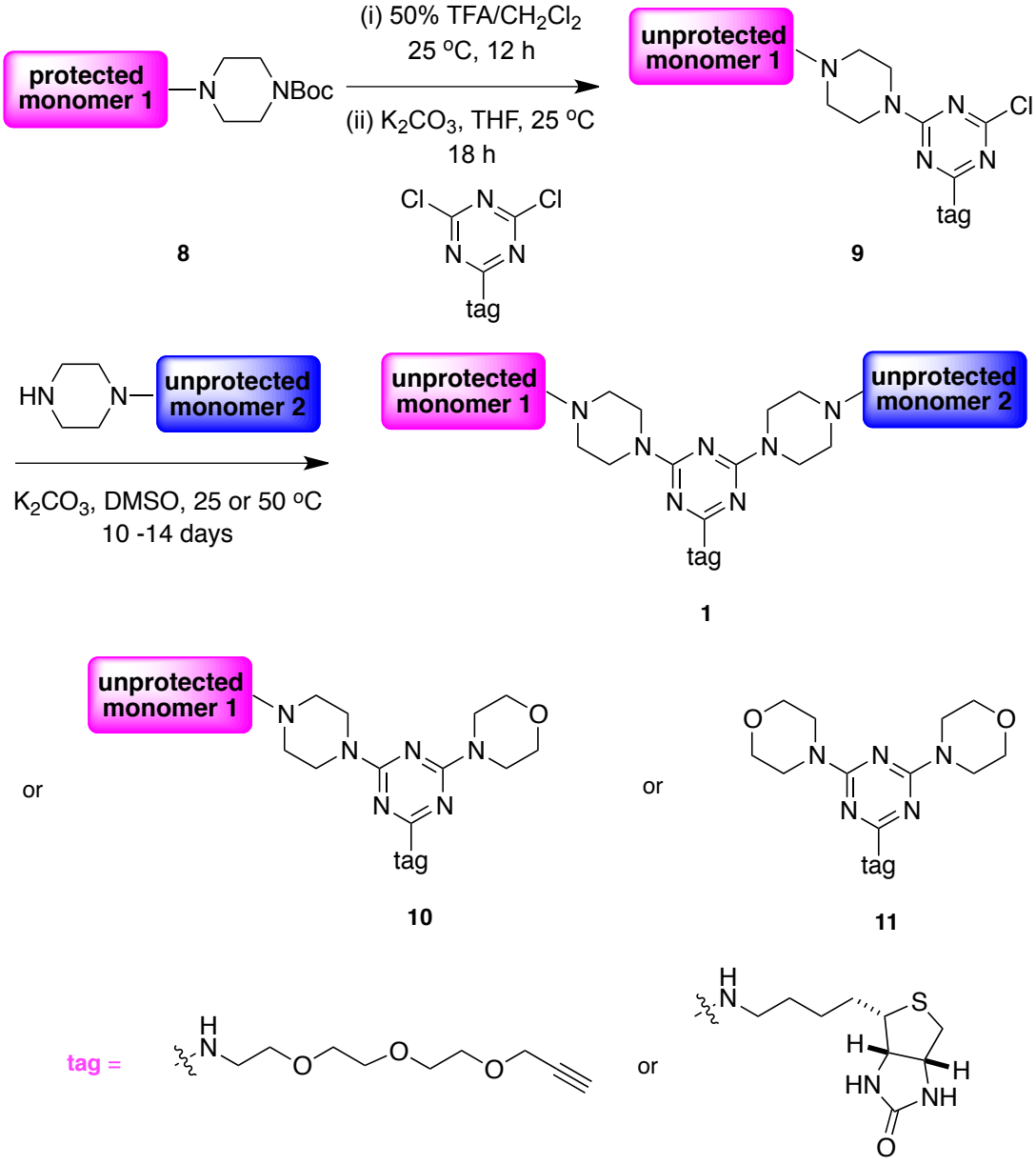
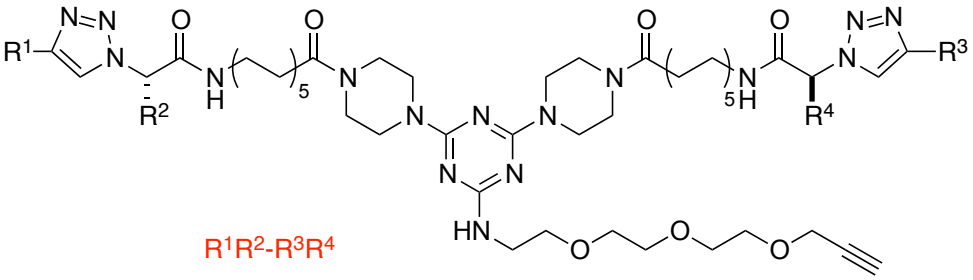
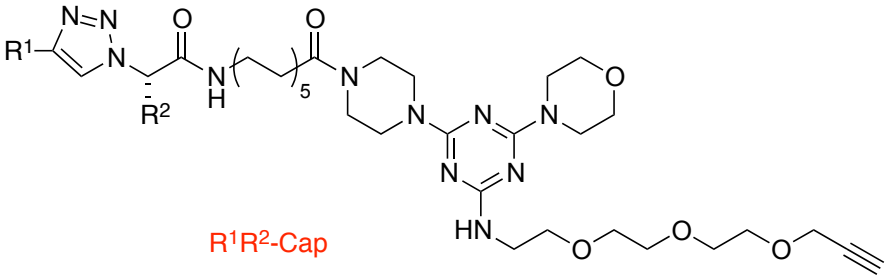


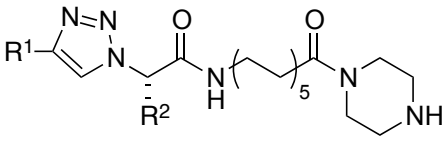
Table S1. Characterization of Compounds with TEG.



R^1R^2-R^3R^4



R^1R^2-Cap



R^1R^2

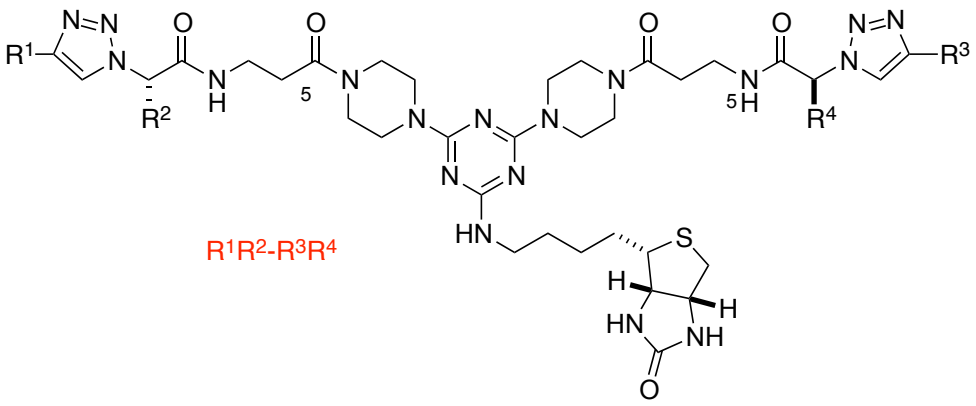
	compound code	sequence (R^1R^2-R^3R^4)	SEDEX purity (%)	retention time (min)	chemical formula	$[M+H]^+$ calculated	$[M+H]^+$ found
1	**KB1365**	KI-KI	100	19.0	$C_{68}H_{121}N_{18}O_7$	1302.0	1302.0
2	**KB1366**	KI-KG	100	17.0	$C_{64}H_{113}N_{18}O_7$	1245.7	1245.9
3	**KB1367**	KI-KE	100	17.2	$C_{67}H_{117}N_{18}O_9$	1317.9	1317.9
4	**KB1368**	KI-KS	100	16.6	$C_{65}H_{115}N_{18}O_8$	1275.9	1275.8
5	**KB1369**	KI-GT	100	16.7	$C_{62}H_{108}N_{17}O_8$	1218.9	1218.9
6	**KB1370**	KI-RI	100	19.6	$C_{68}H_{121}N_{20}O_7$	1330.0	1330.0
7	**KB1371**	KI-SM	100	17.0	$C_{64}H_{112}N_{17}O_8S$	1278.9	1278.9
8	**KB1372**	KG-KG	100	15.6	$C_{60}H_{105}N_{18}O_7$	1189.8	1189.8
9	**KB1373**	KG-KE	100	14.4	$C_{63}H_{109}N_{18}O_9$	1261.9	1262.0
10	**KB1374**	KG-KS	100	13.2	$C_{61}H_{107}N_{18}O_8$	1219.9	1219.8
11	**KB1375**	KG-GT	100	14.8	$C_{58}H_{100}N_{17}O_8$	1162.8	1162.7
12	**KB1376**	KG-RI	100	17.6	$C_{64}H_{113}N_{20}O_7$	1273.9	1273.8
13	**KB1377**	KG-SM	100	16.9	$C_{60}H_{104}N_{17}O_8S$	1222.8	1222.8
14	**KB1378**	KE-KE	100	15.0	$C_{66}H_{113}N_{18}O_{11}$	1333.9	1333.9
15	**KB1379**	KE-KS	91	14.9	$C_{64}H_{111}N_{18}O_{10}$	1291.9	1291.9

16	**KB1380**	KE-GT	100	16.7	$C_{61}H_{104}N_{17}O_{10}$	1234.8	1234.7
17	**KB1381**	KE-RI	100	17.1	$C_{67}H_{117}N_{20}O_9$	1345.9	1345.8
18	**KB1382**	KE-SM	100	16.7	$C_{63}H_{108}N_{17}O_{10}S$	1294.8	1294.8
19	**KB1383**	KS-KS	100	14.7	$C_{62}H_{109}N_{18}O_9$	1249.9	1249.9
20	**KB1384**	KS-GT	100	14.4	$C_{59}H_{102}N_{17}O_9$	1192.8	1192.8
21	**KB1385**	KS-RI	100	17.3	$C_{65}H_{115}N_{20}O_8$	1303.9	1303.9
22	**KB1386**	KS-SM	100	17.0	$C_{61}H_{106}N_{17}O_9S$	1252.8	1252.8
23	**KB1387**	GT-GT	100	15.3	$C_{56}H_{95}N_{16}O_9$	1135.7	1135.6
24	**KB1388**	GT-RI	100	18.8	$C_{62}H_{108}N_{19}O_8$	1246.9	1246.9
25	**KB1389**	GT-SM	98	18.1	$C_{58}H_{99}N_{16}O_9S$	1195.8	1195.6
26	**KB1390**	RI-RI	100	18.9	$C_{68}H_{121}N_{22}O_7$	1358.0	1357.9
27	**KB1391**	RI-SM	100	19.7	$C_{64}H_{112}N_{19}O_8S$	1306.9	1306.9
28	**KB1392**	SM-SM	100	18.8	$C_{60}H_{103}N_{16}O_9S_2$	1255.8	1255.7
29	**KB1393**	SM	100	15.7	$C_{24}H_{45}N_6O_3S$	497.3	497.1
30	**KB1394**	RI	100	11.3	$C_{28}H_{54}N_9O_2$	548.4	548.3
31	**KB1395**	KI	100	11.5	$C_{28}H_{54}N_7O_2$	520.4	520.2
32	**KB1396**	KG	100	10.0	$C_{24}H_{46}N_7O_2$	464.4	464.1
33	**KB1397**	KE	100	10.1	$C_{27}H_{50}N_7O_4$	536.4	536.2
34	**KB1398**	KS	100	8.6	$C_{25}H_{48}N_7O_3$	494.4	494.1
35	**KB1399**	GT	100	9.7	$C_{22}H_{41}N_6O_3$	437.3	437.1
36	**KB1445**	TG-TG	98	14.5	$C_{56}H_{95}N_{16}O_9$	1135.8	1135.5
37	**KB1446**	TG-MS	100	15.8	$C_{58}H_{99}N_{16}O_9S$	1195.8	1195.6
38	**KB1447**	TG-EK	96	13.8	$C_{61}H_{104}N_{17}O_{10}$	1234.8	1234.6
39	**KB1448**	TG-GK	100	14.2	$C_{58}H_{100}N_{17}O_8$	1162.8	1162.5
40	**KB1449**	TG-SK	100	14.4	$C_{59}H_{102}N_{17}O_9$	1192.8	1192.8
41	**KB1450**	TG-IK	100	16.3	$C_{62}H_{108}N_{17}O_8$	1218.9	1218.6
42	**KB1451**	TG-IR	94	17.1	$C_{62}H_{108}N_{19}O_8$	1246.9	1246.7
43	**KB1452**	MS-MS	90	17.2	$C_{60}H_{103}N_{16}O_9S_2$	1255.8	1255.8
44	**KB1453**	MS-EK	93	14.9	$C_{63}H_{108}N_{17}O_{10}S$	1294.8	1294.8
45	**KB1454**	MS-GK	100	16.8	$C_{60}H_{104}N_{17}O_8S$	1222.8	1222.6
46	**KB1455**	MS-SK	100	14.9	$C_{61}H_{106}N_{17}O_9S$	1252.8	1252.7
47	**KB1456**	MS-IK	92	17.7	$C_{64}H_{112}N_{17}O_8S$	1278.9	1278.7
48	**KB1457**	MS-IR	100	17.8	$C_{64}H_{112}N_{19}O_8S$	1306.9	1306.7
49	**KB1458**	EK-EK	100	12.9	$C_{66}H_{113}N_{18}O_{11}$	1333.9	1333.8
50	**KB1459**	EK-GK	97	13.0	$C_{63}H_{109}N_{18}O_9$	1261.9	1261.9
51	**KB1460**	EK-SK	100	12.5	$C_{64}H_{111}N_{18}O_{10}$	1291.9	1292.0
52	**KB1461**	EK-IK	84	15.7	$C_{67}H_{117}N_{18}O_9$	1317.9	1318.0
53	**KB1462**	EK-IR	98	17.0	$C_{67}H_{117}N_{20}O_9$	1345.9	1346.0
54	**KB1463**	GK-GK	100	13.4	$C_{60}H_{105}N_{18}O_7$	1189.8	1190.1
55	**KB1464**	GK-SK	95	13.5	$C_{61}H_{106}N_{18}O_8$	1218.8	1218.7
56	**KB1465**	GK-IK	99	15.4	$C_{64}H_{113}N_{18}O_7$	1245.9	1245.7
57	**KB1466**	GK-IR	96	15.6	$C_{64}H_{113}N_{20}O_7$	1273.9	1273.8
58	**KB1467**	SK-SK	100	12.9	$C_{62}H_{109}N_{18}O_9$	1249.9	1249.7
59	**KB1468**	SK-IK	100	14.8	$C_{65}H_{115}N_{18}O_8$	1275.9	1275.7
60	**KB1469**	SK-IR	89	15.4	$C_{65}H_{115}N_{20}O_8$	1303.9	1303.7
61	**KB1470**	IK-IK	100	16.5	$C_{68}H_{121}N_{18}O_7$	1302.0	1302.0

62	**KB1471**	IK-IR	100	17.8	$C_{68}H_{121}N_{20}O_7$	1330.0	1329.8
63	**KB1472**	IR-IR	100	17.6	$C_{68}H_{121}N_{22}O_7$	1358.0	1357.7
64	**KB1473**	TG	100	8.9	$C_{22}H_{41}N_6O_3$	437.3	437.1
65	**KB1474**	MS	97	11.1	$C_{24}H_{45}N_6O_3S$	497.3	497.2
66	**KB1475**	EK	100	8.6	$C_{27}H_{50}N_7O_4$	536.4	536.3
67	**KB1476**	GK	100	8.8	$C_{24}H_{46}N_7O_2$	464.4	464.2
68	**KB1477**	SK	94	8.0	$C_{25}H_{48}N_7O_3$	494.4	494.2
69	**KB1478**	IK	94	11.4	$C_{28}H_{54}N_7O_2$	520.8	520.3
70	**KB1479**	IR	100	11.4	$C_{28}H_{54}N_9O_2$	548.4	548.4
71	**KB1480**	TG-Cap	100	13.6	$C_{38}H_{64}N_{11}O_7$	786.5	786.3
72	**KB1481**	MS-Cap	100	15.8	$C_{40}H_{68}N_{11}O_7S$	846.5	845.3
73	**KB1482**	EK-Cap	100	10.9	$C_{43}H_{73}N_{12}O_8$	885.6	885.2
74	**KB1483**	GK-Cap	100	14.2	$C_{40}H_{69}N_{12}O_6$	813.5	813.3
75	**KB1484**	SK-Cap	99	12.3	$C_{41}H_{71}N_{12}O_7$	843.6	843.3
76	**KB1485**	IK-Cap	100	16.1	$C_{44}H_{77}N_{12}O_6$	869.6	869.4
77	**KB1486**	IR-Cap	100	18.0	$C_{44}H_{77}N_{14}O_6$	897.6	897.4
78	**KB1551**	KI-IK	100	17.5	$C_{68}H_{121}N_{18}O_7$	1302.0	1301.8
79	**KB1552**	KI-GK	100	15.6	$C_{64}H_{112}N_{18}NaO_7$	1267.9	1267.8
80	**KB1553**	KI-EK	100	16.9	$C_{67}H_{117}N_{18}O_9$	1317.9	1318.1
81	**KB1554**	KI-SK	100	15.1	$C_{65}H_{115}N_{18}O_8$	1275.9	1275.6
82	**KB1555**	KI-TG	100	16.9	$C_{62}H_{108}N_{17}O_8$	1218.9	1219.0
83	**KB1556**	KI-MS	100	18.7	$C_{64}H_{112}N_{17}O_8S$	1278.9	1278.9
84	**KB1557**	KG-IK	100	17.0	$C_{64}H_{113}N_{18}O_7$	1245.9	1246.0
85	**KB1558**	KG-GK	100	14.7	$C_{60}H_{104}N_{18}NaO_7$	1211.8	1211.8
86	**KB1559**	KG-EK	100	14.6	$C_{63}H_{109}N_{18}O_9$	1261.9	1261.9
87	**KB1560**	KG-SK	100	14.1	$C_{61}H_{107}N_{18}O_8$	1219.9	1219.8
88	**KB1561**	KG-TG	100	14.1	$C_{58}H_{100}N_{17}O_8$	1162.8	1162.9
89	**KB1562**	KG-IR	100	18.0	$C_{64}H_{113}N_{20}O_7$	1273.9	1274.1
90	**KB1563**	KG-MS	100	16.6	$C_{60}H_{104}N_{17}O_8S$	1222.8	1222.6
91	**KB1564**	KE-IK	100	16.4	$C_{67}H_{117}N_{18}O_9$	1317.9	1318.1
92	**KB1565**	KE-GK	99	13.5	$C_{63}H_{109}N_{18}O_9$	1261.9	1261.8
93	**KB1566**	KE-EK	100	15.0	$C_{66}H_{113}N_{18}O_{11}$	1333.9	1333.9
94	**KB1567**	KE-SK	100	13.2	$C_{64}H_{110}N_{18}NaO_{10}$	1313.9	1313.9
95	**KB1568**	KE-TG	87	14.3	$C_{61}H_{104}N_{17}O_{10}$	1234.8	1234.9
96	**KB1569**	KE-IR	99	16.7	$C_{67}H_{117}N_{20}O_9$	1345.9	1346.0
97	**KB1570**	KE-MS	100	15.4	$C_{63}H_{108}N_{17}O_{10}S$	1294.8	1294.6
98	**KB1571**	KS-IK	100	16.0	$C_{65}H_{115}N_{18}O_8$	1275.9	1275.8
99	**KB1572**	KS-GK	97	13.6	$C_{61}H_{107}N_{18}O_8$	1219.9	1219.9
100	**KB1573**	KS-EK	100	14.9	$C_{64}H_{111}N_{18}O_{10}$	1291.9	1291.9
101	**KB1574**	KS-SK	100	14.9	$C_{62}H_{109}N_{18}O_9$	1249.9	1249.8
102	**KB1575**	KS-TG	100	14.6	$C_{59}H_{102}N_{17}O_9$	1192.8	1192.7
103	**KB1576**	KS-MS	99	15.4	$C_{61}H_{106}N_{17}O_9S$	1252.8	1252.5
104	**KB1577**	GT-IK	100	17.6	$C_{62}H_{108}N_{17}O_8$	1218.9	1218.7
105	**KB1578**	GT-GK	100	14.4	$C_{58}H_{100}N_{17}O_8$	1162.8	1162.9
106	**KB1579**	GT-EK	100	16.2	$C_{61}H_{104}N_{17}O_{10}$	1234.8	1234.8
107	**KB1580**	GT-SK	100	14.4	$C_{59}H_{102}N_{17}O_9$	1192.8	1192.9

108	**KB1581**	GT-TG	100	14.9	$C_{56}H_{95}N_{16}O_9$	1135.7	1135.9
109	**KB1582**	GT-IR	100	17.1	$C_{62}H_{108}N_{19}O_8$	1246.9	1247.0
110	**KB1583**	GT-MS	100	16.4	$C_{58}H_{99}N_{16}O_9S$	1195.8	1195.7
111	**KB1584**	RI-IK	100	19.3	$C_{68}H_{121}N_{20}O_7$	1330.0	1330.0
112	**KB1585**	RI-GK	100	16.5	$C_{64}H_{113}N_{20}O_7$	1273.9	1273.9
113	**KB1586**	RI-EK	100	17.3	$C_{67}H_{117}N_{20}O_9$	1345.9	1346.1
114	**KB1587**	RI-SK	100	15.7	$C_{65}H_{115}N_{20}O_8$	1303.9	1304.1
115	**KB1588**	RI-TG	100	17.2	$C_{62}H_{108}N_{19}O_8$	1246.9	1246.6
116	**KB1589**	RI-MS	100	18.2	$C_{64}H_{112}N_{19}O_8S$	1306.9	1306.6
117	**KB1590**	SM-IK	100	17.8	$C_{64}H_{112}N_{17}O_8S$	1278.9	1278.7
118	**KB1591**	SM-GK	100	15.7	$C_{60}H_{104}N_{17}O_8S$	1222.8	1222.8
119	**KB1592**	SM-EK	100	14.6	$C_{63}H_{108}N_{17}O_{10}S$	1294.8	1294.9
120	**KB1593**	SM-SK	100	15.0	$C_{61}H_{106}N_{17}O_9S$	1252.8	1252.6
121	**KB1594**	SM-TG	100	15.6	$C_{58}H_{99}N_{16}O_9S$	1195.8	1195.6
122	**KB1595**	SM-IR	100	18.3	$C_{64}H_{112}N_{19}O_8S$	1306.9	1306.8
123	**KB1596**	SM-MS	96	15.5	$C_{60}H_{103}N_{16}O_9S_2$	1255.8	1255.6
124	**KB1597**	KI-Cap	100	16.1	$C_{44}H_{77}N_{12}O_6$	869.6	869.6
125	**KB1598**	KG-Cap	100	13.5	$C_{40}H_{69}N_{12}O_6$	813.5	813.4
126	**KB1599**	KE-Cap	94	14.6	$C_{43}H_{73}N_{12}O_8$	885.6	885.5
127	**KB1600**	KS-Cap	100	13.0	$C_{41}H_{71}N_{12}O_7$	843.6	843.5
128	**KB1601**	GT-Cap	100	16.6	$C_{38}H_{64}N_{11}O_7$	786.5	786.5
129	**KB1602**	RI-Cap	100	15.9	$C_{44}H_{77}N_{14}O_6$	897.6	897.6
130	**KB1603**	SM-Cap	100	14.8	$C_{40}H_{68}N_{11}O_7S$	846.5	846.3
131	**KB1604**	KI-IR	100	18.9	$C_{68}H_{121}N_{20}O_7$	1330.0	1329.9
132	**KB1605**	KS-IR	100	15.3	$C_{65}H_{115}N_{20}O_8$	1303.9	1303.9
133	**KB1606**	RI-IR	100	19.6	$C_{68}H_{121}N_{22}O_7$	1358.0	1357.7

Table S2. Characterization of Compounds with Biotin.



	code	TEG code	seq.[a] (R^1R^2-R^3R^4)	SEDEX purity (%)	time[b] (min)	chemical formula	$[M+H]^+$ calculated	$[M+H]^+$ found
1	**KB1923**	KB1468	SK-IK	100	13.2	$C_{65}H_{115}N_{20}O_6S$	1303.9	1303.6
2	**KB1924**	KB1471	IK-IR	100	14.8	$C_{68}H_{121}N_{22}O_5S$	1358.0	1358.0
3	**KB1925**	KB1579	GT-EK	100	13.2	$C_{61}H_{104}N_{19}O_8S$	1262.8	1262.8
4	**KB1926**	KB1588	RI-TG	100	14.2	$C_{62}H_{108}N_{21}O_6S$	1274.9	1274.8
5	**KB1927**	KB1591	SM-GK	100	13.8	$C_{60}H_{104}N_{19}O_6S_2$	1250.7	1250.7
6	**KB1811**	KB1368	KI-KS	100	13.1	$C_{65}H_{116}N_{20}O_6S$	1304.9	1304.8

[a] seq. = sequence. [b] time = retention time.

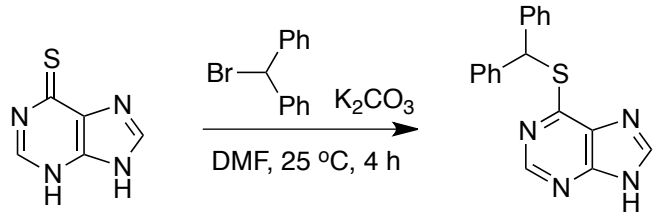
APPENDIX C

EXPERIMENTAL FOR CHAPTER III

A. Syntheses For 6-Mercaptopurine Derivative 3

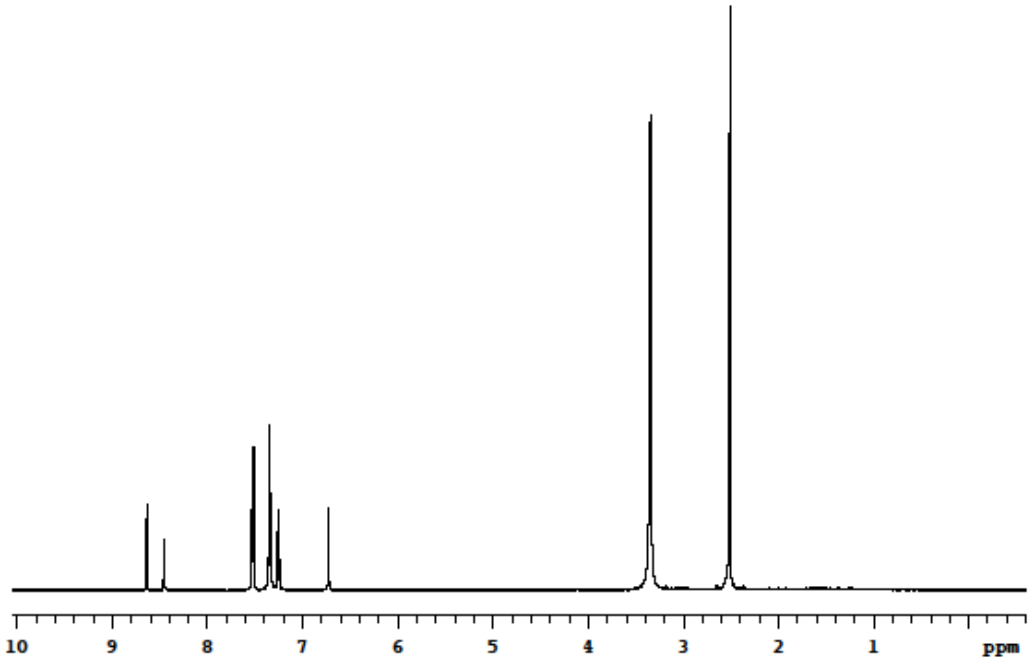
General procedure for Compound 1

To a solution of mercaptopurine (9.9 mmol, 1.0 equiv) in DMF (1.0 M) was added K_2CO_3 (9.9 mmol, 1.0 equiv) under N_2 (g). The solution was stirred at 25 °C for 15 min. To the mixture was added bromodiphenylmethane (9.9 mmol, 1.0 equiv), and the mixture was stirred at 25 °C for 4 h. The mixture was diluted with water (90 mL), and then acidified with acetic acid. The mixture was stirred vigorously for 30 min, filtered, and then dried to give an amorphous solid. Recrystallization from MeOH gave 2.1 g (67 %) of compound **1** as white powder.



1

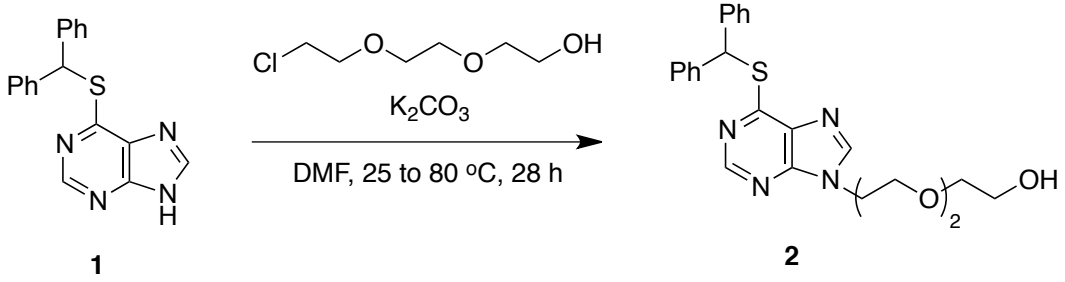
^{1}H NMR (500 MHz, DMSO) δ 8.63 (s, 1H), 8.45 (s, 1H), 7.51 (d, 4H, J = 7.5 Hz), 7.33 (t, 4H, J = 7.5 Hz), 7.24 (7, 2H, J = 7.5 Hz), 6.72 (s, 1H)



^{1}H NMR of **1**

General procedure for Compound 2

To a solution of compound **1** (3.1 mmol, 1.0 equiv) in DMF (0.5 M) was added K_2CO_3 (3.1 mmol, 1.0 equiv) and 2-[2-9chloroethoxyl)-ethoxy]ethanol (3.1 mmol, 1.0 equiv) under N_2 (g) at 25 °C. The solution was warmed to 80 °C and then stirred at the temperature for 2 h. DMF was removed by lyophilizer, and then the mixture was purified by flash chromatography (1:19 MeOH/CH_2Cl_2). 1.2 g (82 %) of compound **2** was obtained as colorless oil.

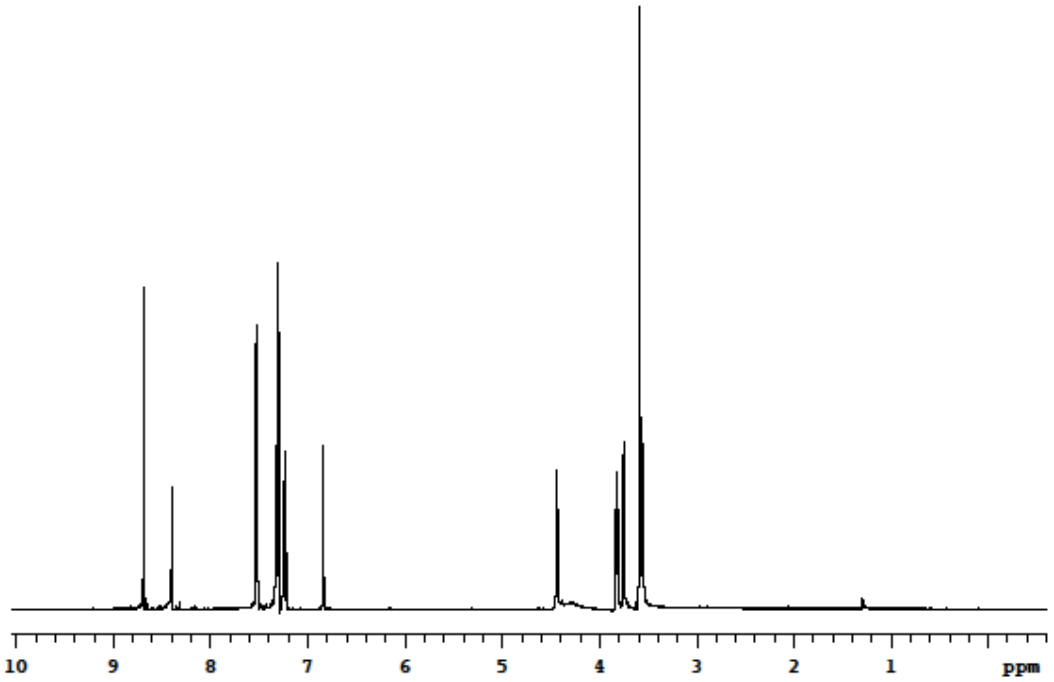


1 **2**

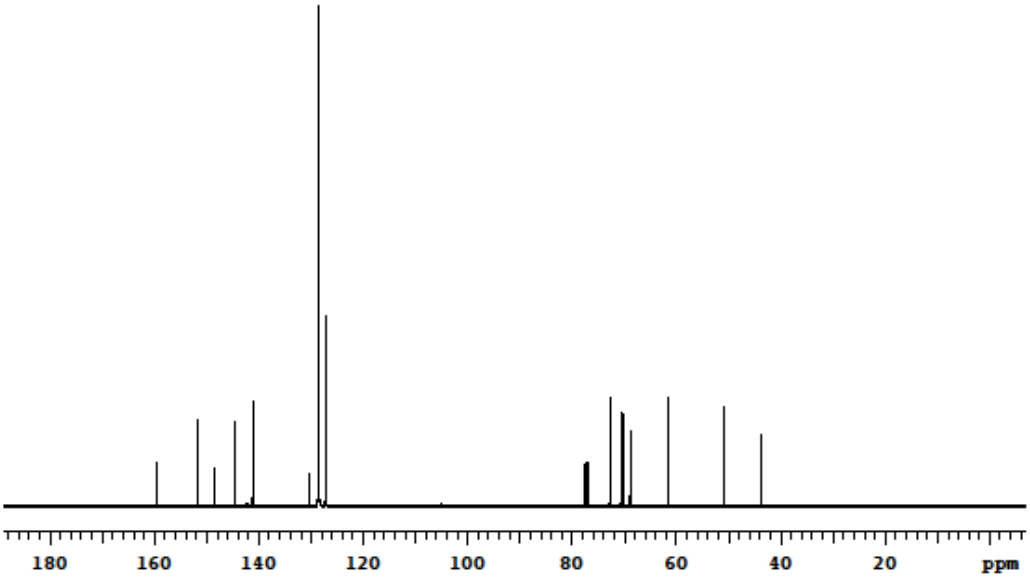
^{1}H NMR (500 MHz, $CDCl_3$) δ 8.67 (s, 1H), 8.38 (s, 1H), 7.51 (d, 4H, J = 7.5 Hz), 7.30 (dd, 4H, J = 7.0, 13.5 Hz), 7.22 (t, 2H, 7.5 Hz), 6.82 (s, 1H), 4.42 (t, 2H, J = 4.5 Hz), 3.80 (t, 2H, J = 5.0 Hz), 3.73 (t, 2H, J = 4.5 Hz), 3.57 (m, 4H), 3.54 (t, 2H, J = 4.0 Hz)

^{13}C NMR (125 MHz, $CDCl_3$) δ 159.2, 151.5, 148.2, 144.3, 140.9, 130.1, 128.4, 128.2, 127.0, 72.5, 70.3, 69.9, 68.6, 61.3, 50.5, 43.4

MS (ESI, m/z) calcd for $C_{24}H_{27}N_4O_3S$ $(M+H)^+$ 451.1, found 451.2



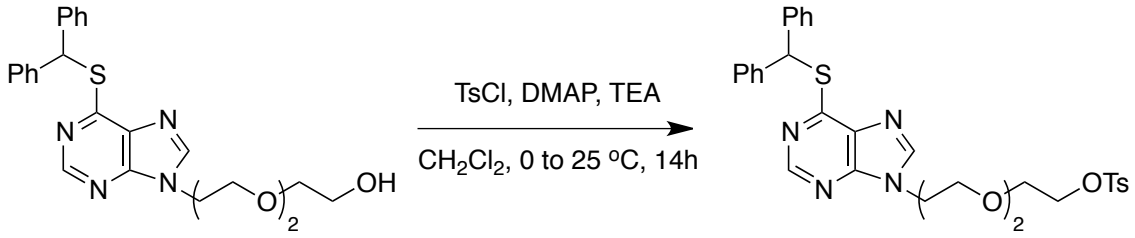
^{1}H NMR of **2**



^{13}C NMR of **2**

General procedure for Compound 3

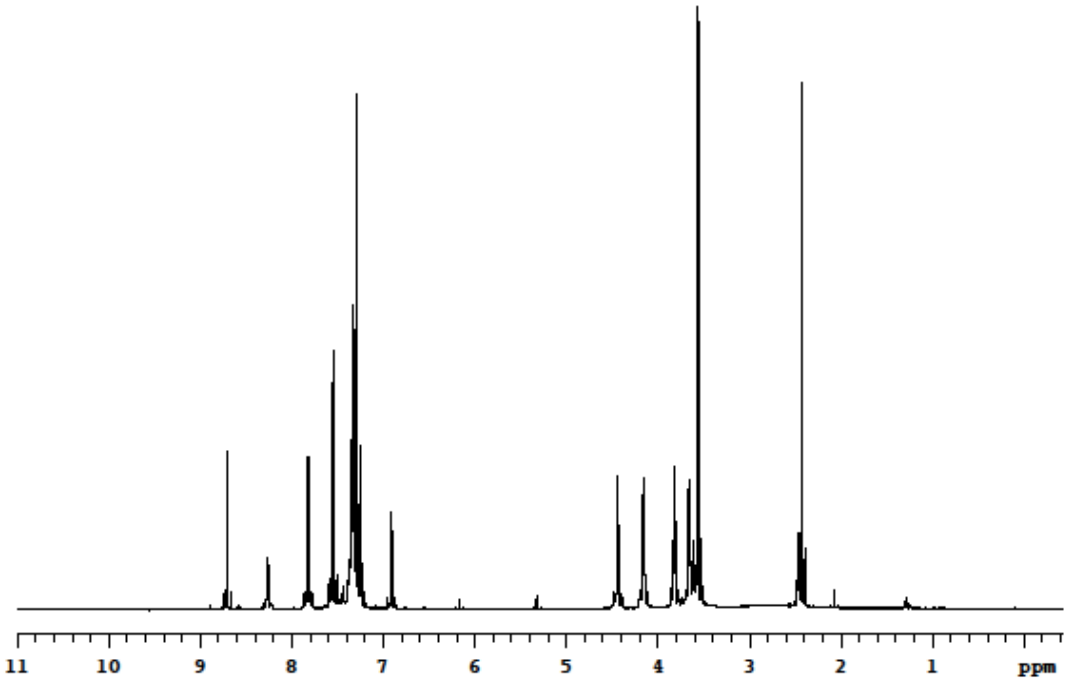
To a solution of compound **2** (2.6 mmol, 1.0 equiv) in CH_2Cl_2 (0.15 M) was added DMAP (0.026 mmol, 0.01 equiv) and TEA (6.4 mmol, 2.5 equiv) under N_2 (g) at 0 °C. 4-Toluenesulfonyl chloride (5.13 mmol, 2.0 equiv) was portioned over 10 min at 0 °C. The solution was stirred at 0 °C for 10 min, and then the solution was warmed to 25 °C. After 2 h, DMAP (0.26 mmol, 0.1 equiv) was added more to the mixture, and then stirred at 25 °C for 14 h. The mixture was diluted with water and then extracted with CH_2Cl_2. The combined organic phases were dried over $MgSO_4$. After completely removing the solvent, tosyl-protected compound was purified by flash chromatography (1:49 MeOH/CH_2Cl_2). 1.3 g (84%) of compound tosyl-protected compound was obtained as colorless oil.



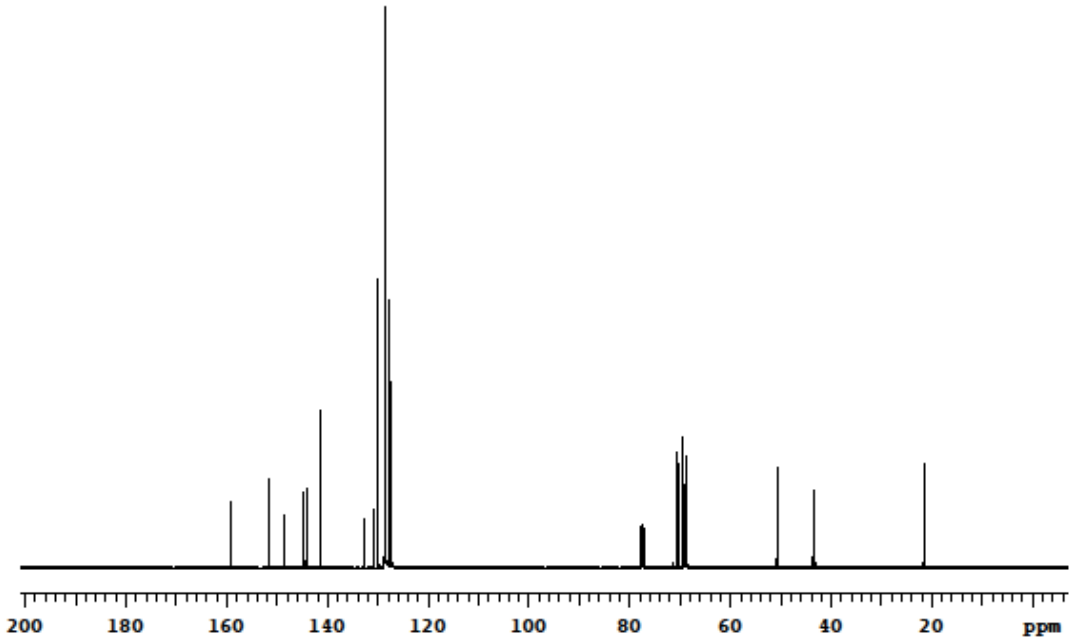
2 **2'**

^{1}H NMR (500 MHz, $CDCl_3$) δ 8.65 (s, 1H), 8.10 (s, 1H), 7.76 (d, 2H, J = 8.5 Hz), 7.52 (d, 4H, J = 8.0 Hz), 7.27 (m, 6H), 7.18 (t, 2H, J = 7.0 Hz), 6.85 (s, 1H), 4.32 (t, 2H, J = 5.0 Hz), 4.08 (t, 2H, J = 3.0 Hz), 3.72 (t, 2H, J = 4.5 Hz), 3.55 (t, 2H, J = 4.5 Hz), 3.46 (m, 4H), 2.34 (s, 3H)

^{13}C NMR (125 MHz, $CDCl_3$) δ 158.8, 151.2, 148.3, 144.5, 143.7, 140.9, 132.4, 130.3, 129.5, 128.2, 128.1, 127.5, 126.8, 70.2, 69.9, 68.9, 68,6, 68.2, 50.3, 43.2, 21.2

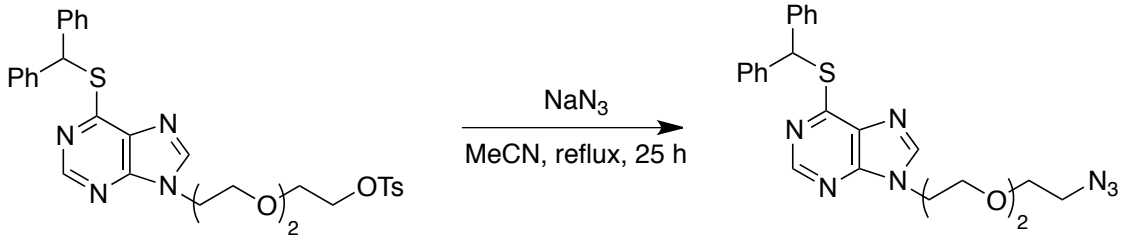


^{1}H NMR of **2'**



^{13}C NMR of **2'**

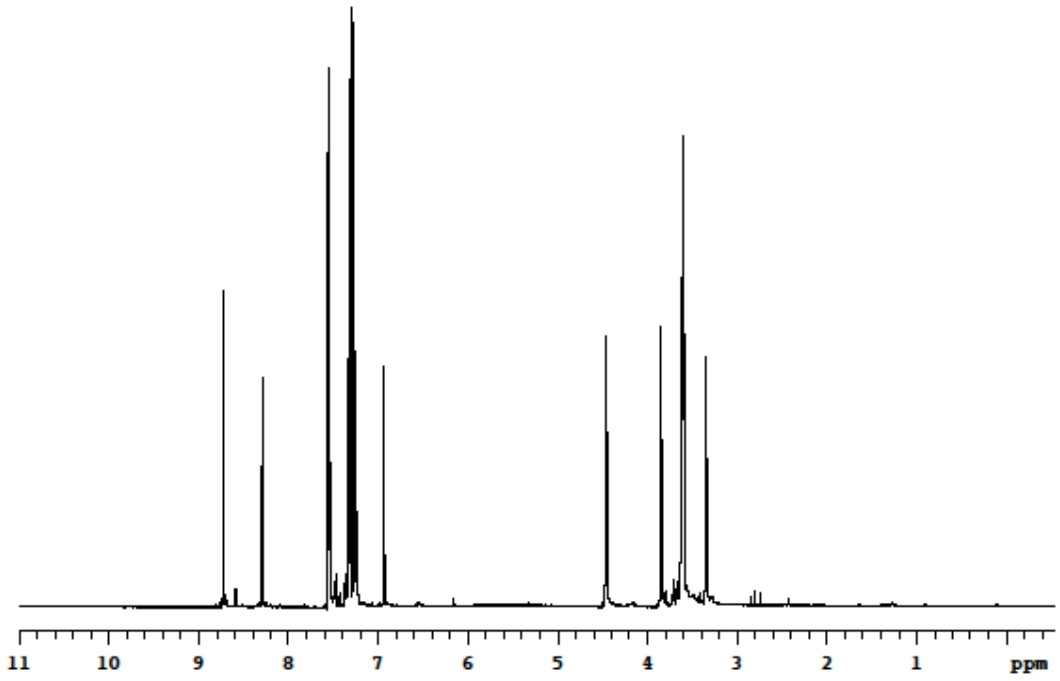
To a solution of compound **2'** (2.1 mmol, 1.0 equiv) in MeCN (0.1 M) was added NaN3 (3.1 mmol, 1.5 equiv) under N_2 (g). The mixture was refluxed for 25 h. After 25 h, the mixture was cooled down to room temperature, and diluted with water, and then extracted with CH_2Cl_2. The combined organic phases were dried over $MgSO_4$. After completely removing the solvent, compound **3** was purified by flash chromatography (3:97 MeOH/CH_2Cl_2). 1.0 g (97%) of compound **3** was obtained as yellowish oil.



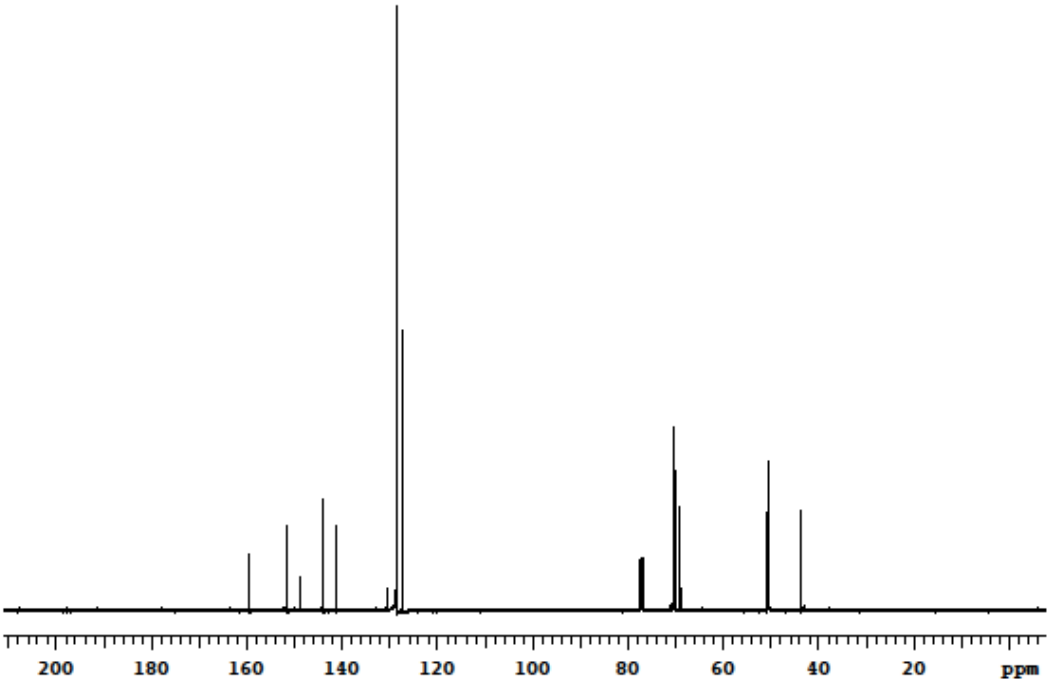
2' **3**

^{1}H NMR (500 MHz, $CDCl_3$) δ 8.70 (s, 1H), 8.28 (s, 1H), 7.53 (d, 4H, J = 8.0 Hz), 7.31 (t, 4H, J = 7.0 Hz), 7.23 (t, 2H, J = 80 Hz), 6.92 (s, 1H), 4.44 (t, 2H, J = 5.0 Hz), 3.82 (t, 2H, J = 5.0 Hz), 3.59 (m, 6H), 3.33 (t, 2H, J = 5.0 Hz)

^{13}C NMR (125 MHz, $CDCl_3$) δ 159.3, 151.5, 148.4, 143.9, 141.0, 130.5, 128.5, 128.3, 127.1, 70.4, 69.9, 68.9, 50.6, 50.4, 43.6



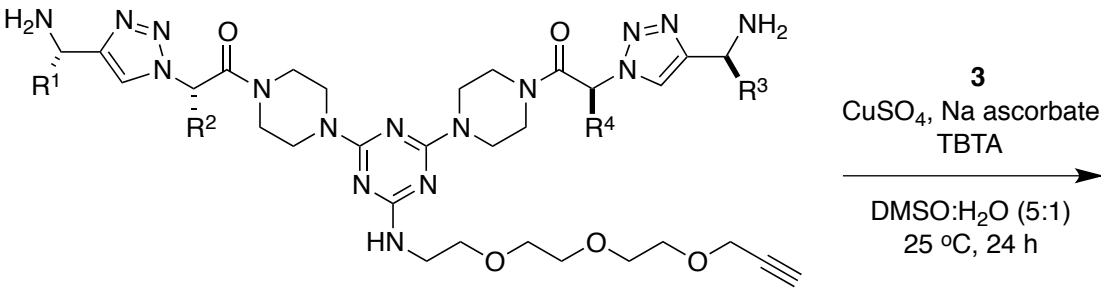
^{1}H NMR of **3**



^{13}C NMR of **3**

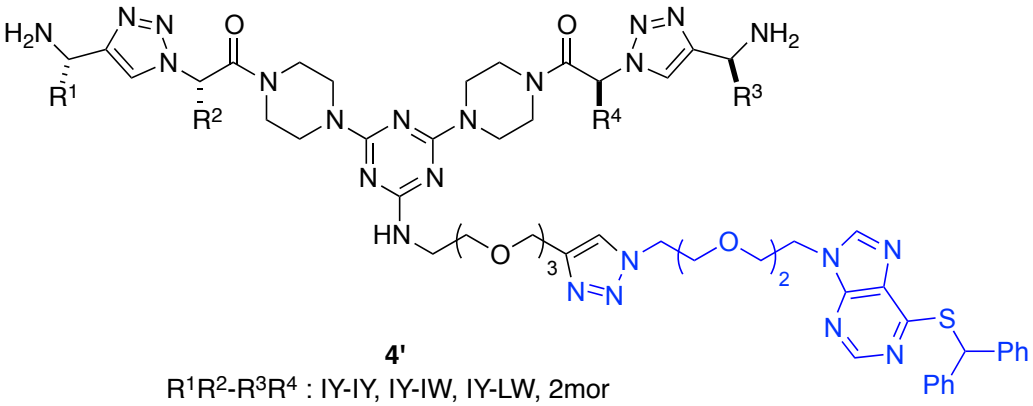
B. Syntheses For Bivalent Mimics Containing 6-Mercaptopurine 5

Compounds **4** (IY-IY-TEG, IY-IW-TEG, IY-LW-TEG, and 2mol-TEG) were prepared via the previously published methodology in our group. To a solution of compound **3** (1.0 equiv) and compound **4** (1.0 equiv) in DMSO:H_2O (5:1, 0.01 M) were added TBTA (0.01 equiv) at 25 °C. $CuSO_4$ (0.1 equiv, from 0.05 M aqueous solution) and fresh Na ascorbate (0.4 equiv, from 0.05 M aqueous solution) were added to the mixture at 25 °C. The mixture was stirred at 25 °C for 24 h. The reactions were monitored by analytical HPLC. The crude compounds were lyophilized to remove DMSO, and then purified with RP-reparative HPLC to obtain the bivalent mimics **4'**.



4

R^1R^2-R^3R^4 : IY-IY, IY-IW, IY-LW, 2mol



4'

R^1R^2-R^3R^4 : IY-IY, IY-IW, IY-LW, 2mor

Table S3. Characterization of Compounds **4'**.

com'd	sequence (R^1R^2-R^3R^4)	retention time (min)	chemical formula	$[M+H]^+$ calculated	$[M+H]^+$ found
4'a	IY-IY	14.4	$C_{76}H_{100}N_{23}O_9S$	1510.8	1510.6
4'b	IY-IW	15.0	$C_{78}H_{101}N_{24}O_8S$	1533.8	1533.4
4'c	IY-LW	15.0	$C_{78}H_{101}N_{24}O_8S$	1533.8	1533.9
4'd	2mol	17.2	$C_{44}H_{59}N_{13}O_7S$	913.4	913.9

Deprotection of diphenylmethyl group from thione **4'** was performed with 50 % TFA/CH_2Cl_2 and phenol (3.0 equiv) at 25 °C. The mixture was stirred at 25 °C for 16 h. The reactions were monitored by analytical HPLC. The crude compounds were purified with RP-reparative HPLC to obtain the bivalent mimics **5**. The final products **5** were lyophilized three times in 1.0 % acetic acid to remove TFA.

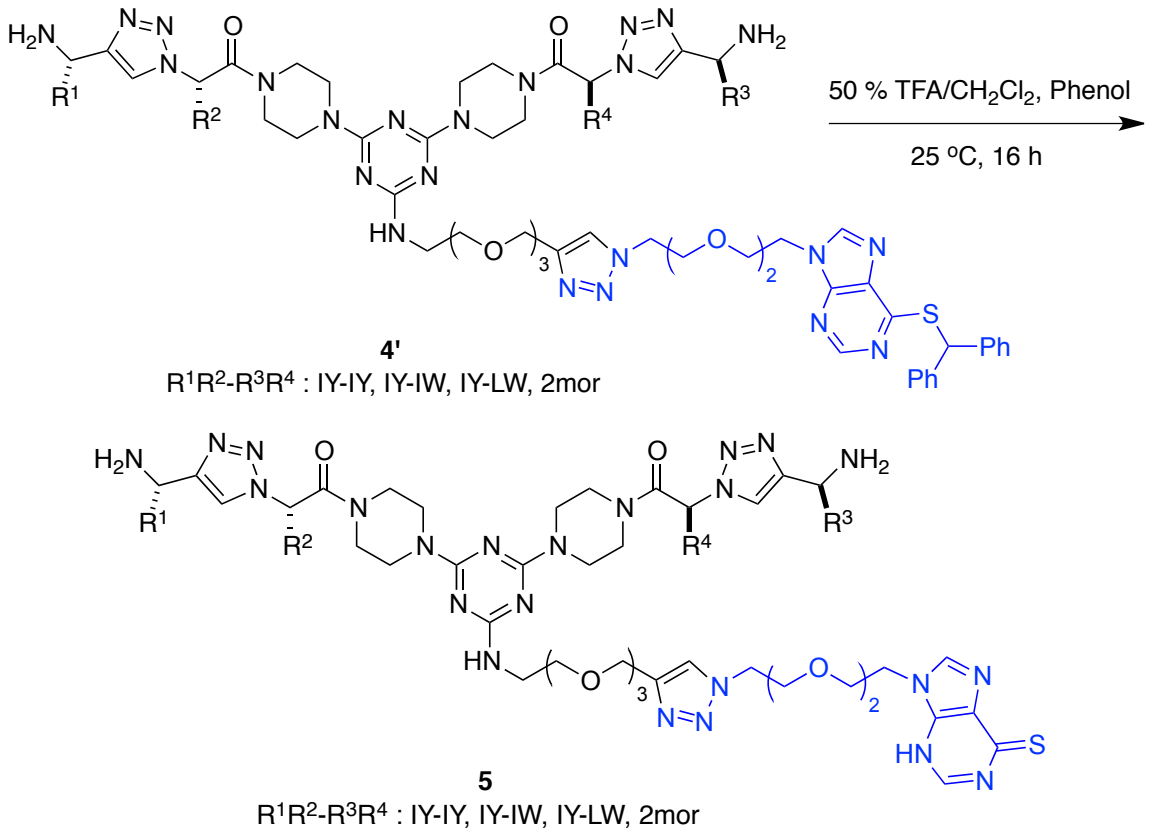


Table S4. Characterization of Compounds **5**.

com'd	sequence (R^1R^2-R^3R^4)	retention time (min)	chemical formula	$[M+H]^+$ calculated	$[M+H]^+$ found
5a	IY-IY	10.6	$C_{63}H_{90}N_{23}O_9S$	1344.7	1345.1
5b	IY-IW	11.4	$C_{65}H_{91}N_{244}O_8S$	1367.7	1367.8
5c	IY-LW	11.6	$C_{65}H_{91}N_{244}O_8S$	1367.7	1367.5
5d	2mol	10.5	$C_{31}H_{49}N_{13}O_7S$	747.4	746.7

C. Antiproliferative Assay

TrkC overexpressed NIH-3T3 cells were kindly provided by Dr David Kaplan at University of Toronto in Canada, and wild-type NIH-3T3 cells were provided by Dr Jean-Philippe Pellois at Texas A&M University. $TrkC^+$ NIH-3T3 cells (TrkC) (10,000 cells/well, 50 μL in Dulbecco's Modified Eagle Medium/nutrient mixture F-12 (DMEM/F12) including G418) were plated on 96-well plates and allowed to adhere at

37 °C in 5 % CO_2 and 95 % air for 3 h. Thereafter, the cells were treated with 50 μL aliquot of each test compounds in PFHM-II and 10 % NBCS (newborn calf serum) at different concentrations, ranging from 0.1 μM to 160 μM. The final concentration of NBCS was 5 %. The cells were then incubated for 60 h. For wild-type NIH-3T3 cells, 50 μL in Dulbecco's Modified Eagle Medium/nutrient mixture F-12 (DMEM/F12) that does not include G418 was used, but other process are the same to TrkC Cells' one.

The cell's viability was assessed through an MTT conversion assay. Briefly, 20 μL of MTT (5mg/mL, in Hank's balanced salt solution, HBSS) were added and the cells were incubated for an additional 2 – 3 h. Thereafter, the cells were lysed and the dark blue crystals solubilized with 100 μL of an aqueous solution containing 35 % (v/v) DMF, 15 % (v/v) glacial acetic acid, 15 % (w/v) SDS with an adjusted pH of 3.8.

The optical density (OD) of each well (at 570 mm) was measured with a BioTek Synergy 4 Microplate Reader. The viability of each cell line in response to the treatment with tested compounds was calculated as: % dead cells = 100 – (OD treated/OD control) × 100.

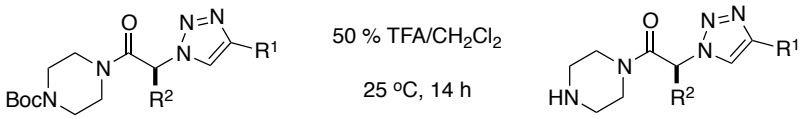
APPENDIX D

EXPERIMENTAL FOR CHAPTER IV

A. General Procedure and Preparation of Bivalent Mimics 10a and 10e

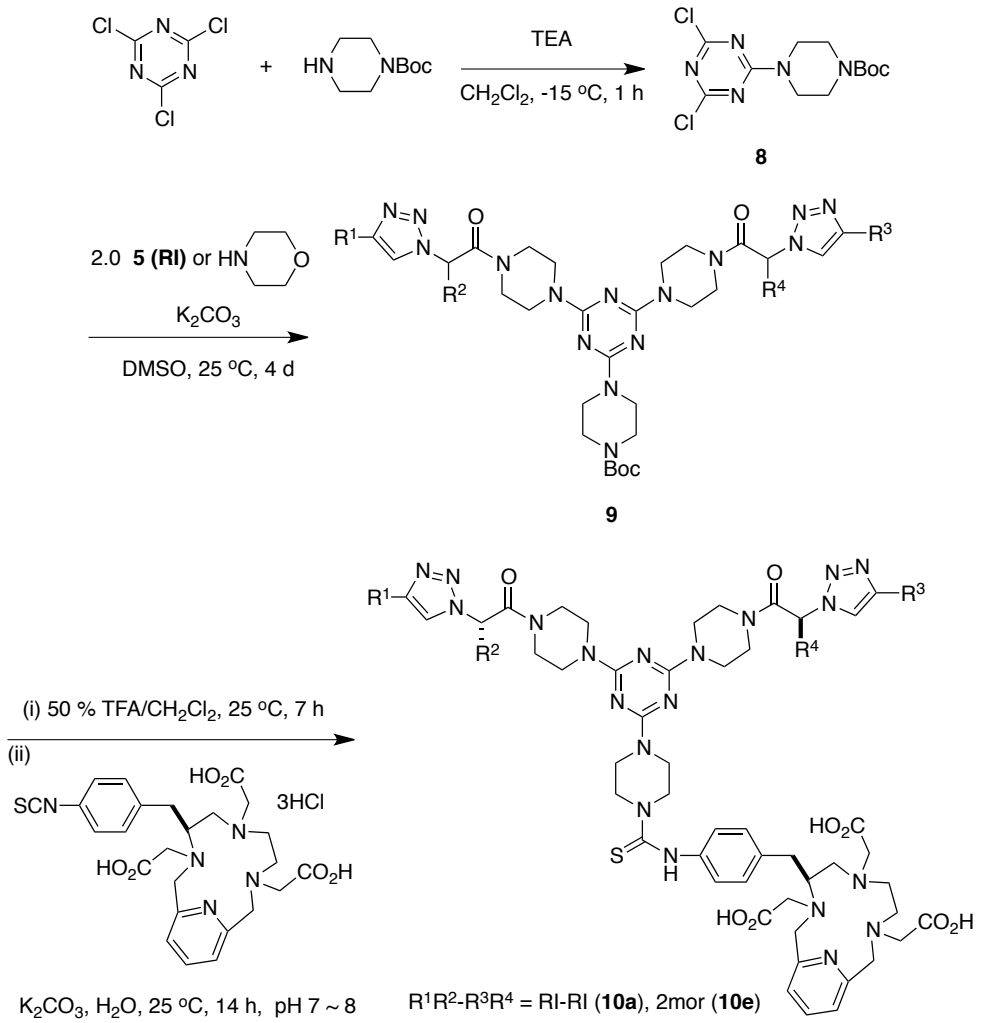
Boc-protected triazole-based monovalent mimics **4** were prepared by Dr Yu Angell. Boc-protecting group in compound **4** was deprotected with 50 % TFA/CH_2Cl_2 at 25 °C for 14 h. To solution of cyanuric chloride (1.0 equiv) and TEA (1.1 equiv) in CH_2Cl_2 (0.1 M) was Boc-piperazine (1.0 equiv) at -15 °C. The mixture was stirred at -15 ^oC for 1 h, and then purified by flash column chromatography to obtain compound **8**. To a solution of compound **5** (2.0 equiv) or morpholine (2.0 equiv) were added K_2CO_4 (4.0 equiv) and DMSO (0.05 M) at 25 °C for 4 d. The reaction was monitored by analytical HPLC. After the reaction was done, the crude compound was lyophilized to remove DMSO, and then purified with RP-preparative HPLC to get a pure compound **9**. Compound **9** was deprotected with 50 % TFA/CH_2Cl_2. After removing the solvent, the deprotected compound was neutralized to pH 7-8 with K_2CO_3 in deionized water, and then p-SCN-Bn-PCTA (0.09 equiv) was added into the solution at 25 °C. pH of the solution was adjusted to pH 7-8 again with K_2CO_3 after adding p-SCN-Bn-PCTA. The reaction was monitored by analytical HPLC, and after the reaction was done, the mixture was purified with RP-preparative HPLC to get a pure compound **10a** or **10e**.

Scheme S1. Preparation of the Tagged Bivalent Peptidomimetics **10a** and **10e**.



4 **5**

R^1R^2 = RI, SY, KT
KG, SG, TG, KI

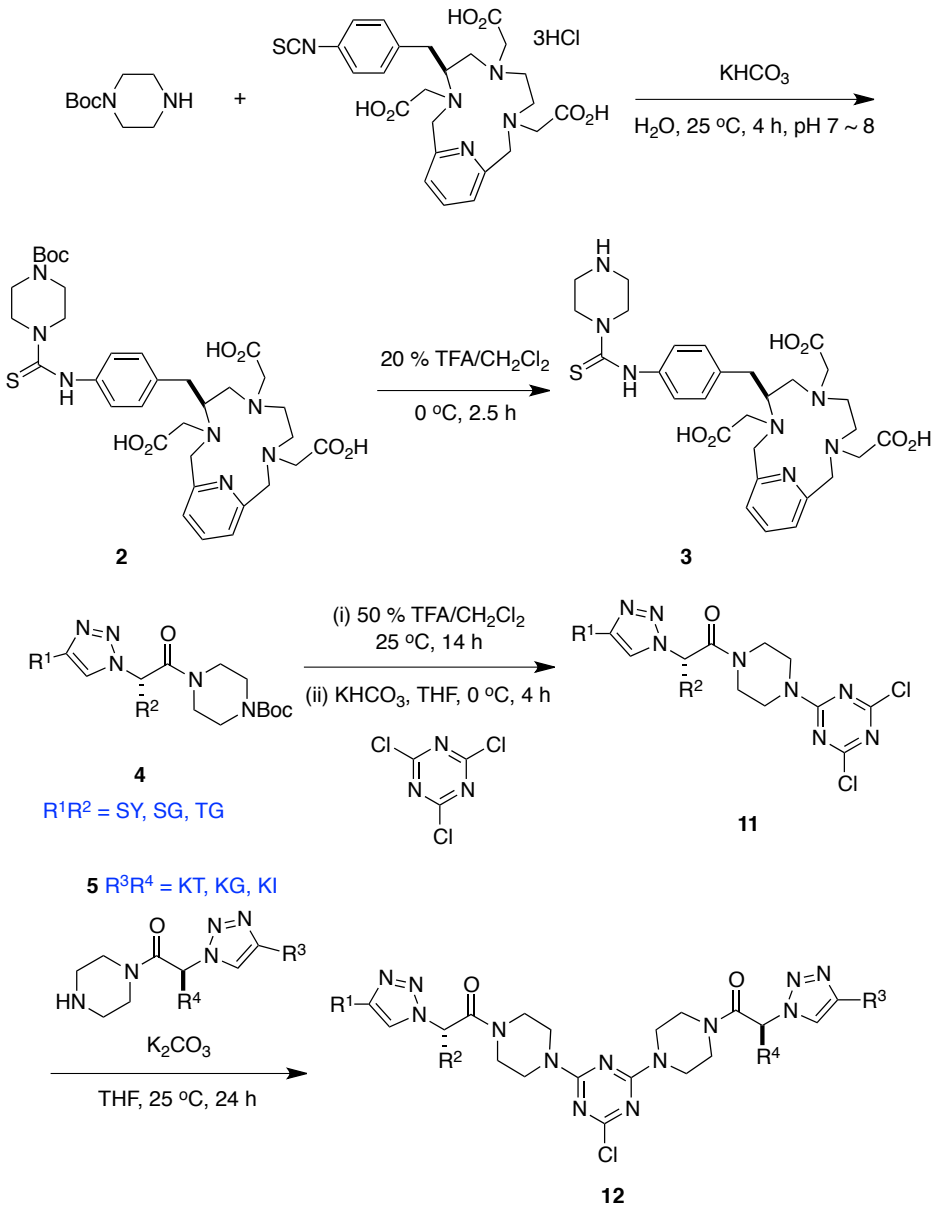


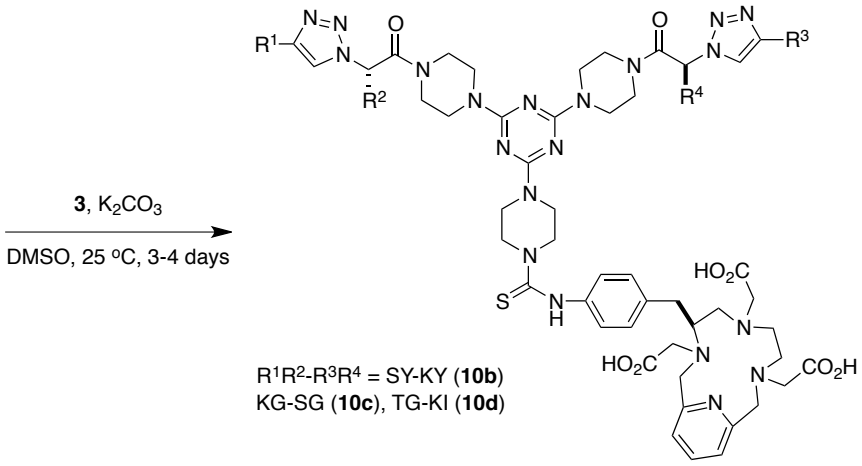
B. General Procedure and Preparation of Bivalent Mimics 10b - d

Boc-protecting group in compounds **4** for SY, SG, and TG was deprotected with 50 % TFA/CH_2Cl_2 at 25 °C for 14 h. To a solution of the deprotected compound (1.0 equiv) in THF (0.05 M) was added $KHCO_3$ (about 4.0 equiv) at 0 °C. The mixture was stirred at 0 °C for 4 h. The reaction was monitored by analytical HPLC. After the reaction was done, the second deprotected monovalent mimics (KT, KG, and KI, 1.0 equiv) were added into the mixture **11**, and the solution was basified with K_2CO_3 (about 4.0 equiv) at 25 °C. The mixture was stirred at 25 °C for 24 h. The reaction was monitored by analytical HPLC, and the compound **12** was purified with RP-preparative HPLC.

Compound **2** was deprotected with 20 % TFA/CH_2Cl_2 at 0 °C for 2.5 h to obtain compound **3**. To a solution of compounds **12** (1.0 equiv) in DMSO (0.05 M) were added K_2CO_3 (4.0 equiv) and compound **3** (1.0 equiv) at 25 °C. The mixture was stirred at 25 °C for 3 ~ 4 d. The reaction was monitored by analytical HPLC, and the compound **10b - d** was purified with RP-preparative HPLC.

Scheme S2. Preparation of the Tagged Bivalent Peptidomimetics **10b** - **d**.

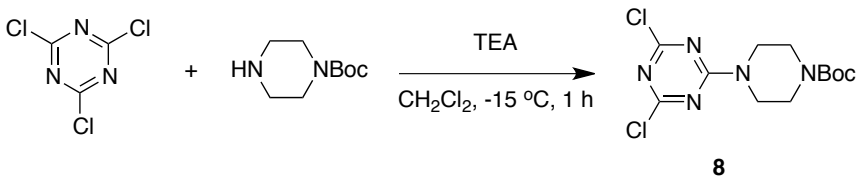




C. Syntheses For Compound 8 and 2

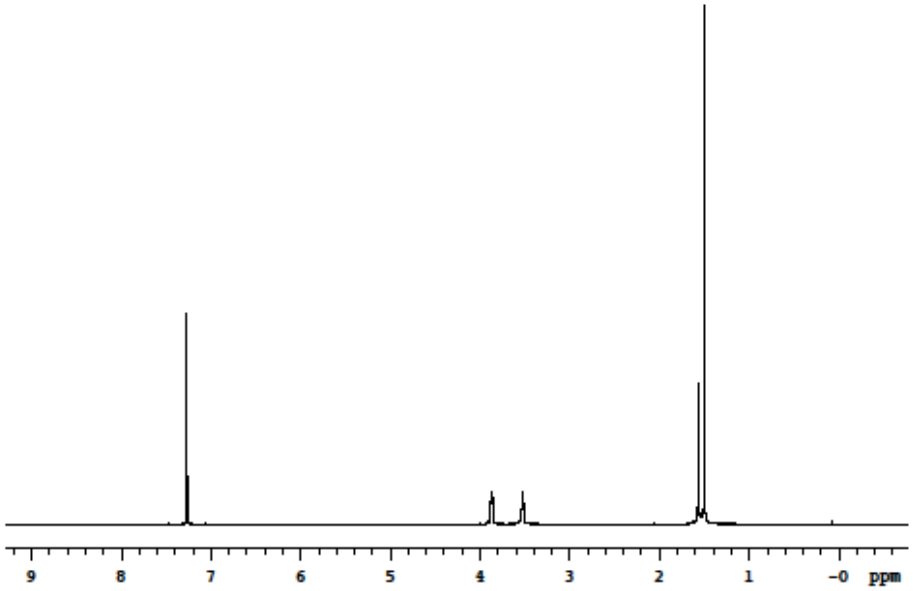
Preparation of Compound 8

To solution of cyanuric chloride (1.0 equiv) and TEA (1.1 equiv) in CH_2Cl_2 (0.1 M) was Boc-piperazine (1.0 equiv) at -15 °C. The mixture was stirred at -15 °C for 1 h, and then purified by flash column chromatography (1:9 EtOAc/Hexanes) to obtain **8** (82 %) as white solid.



8

^{1}H NMR (500 MHz, CD_3OD) δ 3.87 (s, 4H), 3.52 (s, 4H), 1.50 (s, 9H)

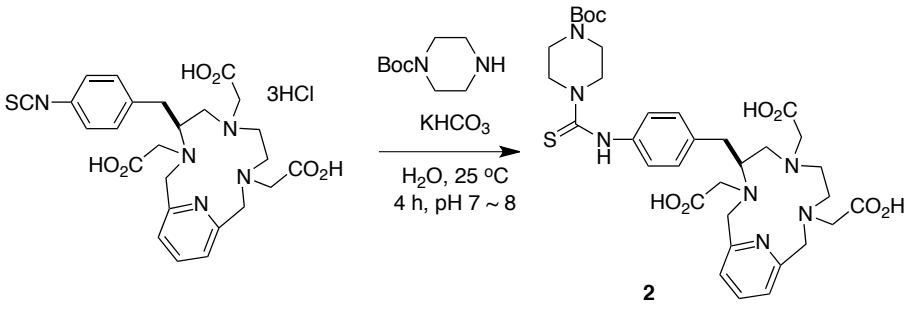


^{1}H NMR of **8**

Preparation of Compound 2

SCN-Bn-PCTA was bought from Macrocyclics.inc.

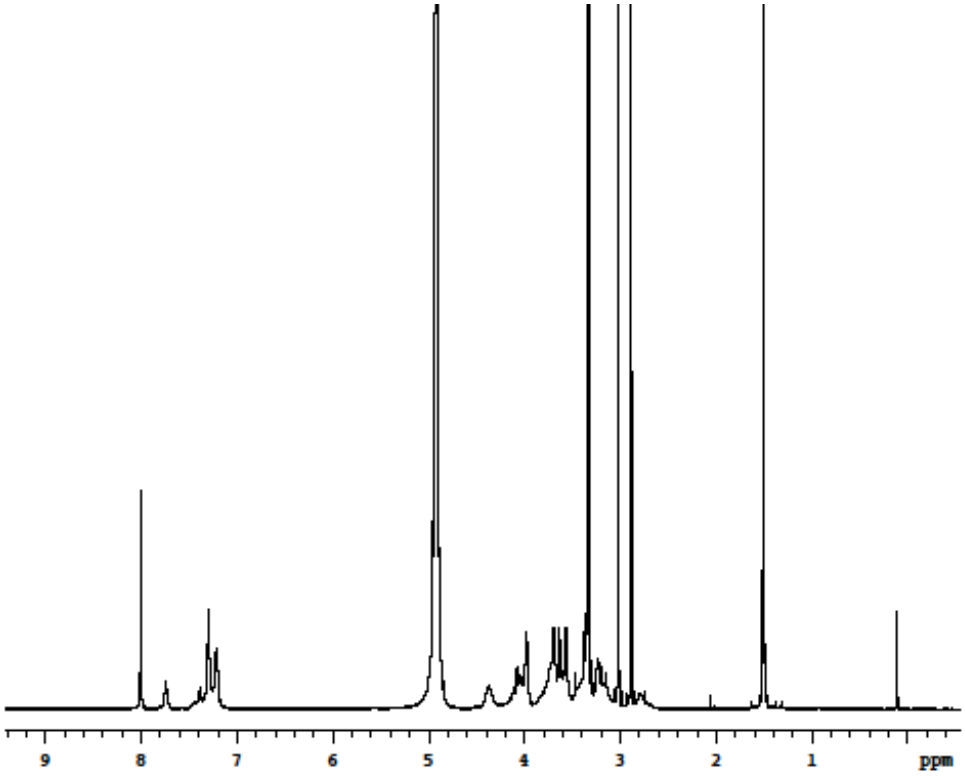
To a solution of SCN-Bn-PCTA•3HCl (1.0 equiv) and Boc-piperazine (1.0 equiv) in H_2O (0.1 M) was added K_2CO_3 (about 4.0 equiv) at 25 °C. The solution was adjusted to pH 7~8 with K_2CO_3. The mixture was stirred at 25 °C for 4 h. The crude compounds were lyophilized to remove H_2O. The compound **2** was used without further purification.



2

^{1}H NMR (500 MHz, CD_3OD) δ 7.74 (s, 1H), 7.38 (s, 1H), 7.30 (s, 3H), 7.20 (s, 2H), 4.36 -4.31 (br, 2H), 4.15-4.00 (m, 8H), 3.69-3.57 (m, 11H), 3.41-3.15 (m, 8H), 1.50 (s, 9H)

MS (ESI, m/z) calcd for $C_{32}H_{48}N_7O_8S$ $(M+H)^+$ 714.3, found 714.3



^{1}H NMR of **2**

D. Syntheses Of Eu-labeled Bivalent Mimics 1a – e

For Eu-binding, the compounds **10a - e** (1.0 equiv) were dissolved in 0.1 M sodium acetate buffer (pH 5.5, 0.05M). $EuCl_3 \cdot 6H_2O$ (3.0 equiv) in the acetate buffer (pH 5.5, 0.1 M) was added to the solution, and the mixture was stirred at 25 ^oC for 24 h. The reaction was monitored by analytical HPLC. After the reaction was done, the mixture was purified with RP-preparative HPLC. 0.1 % acetic acid was used to remove the TFA three times under lyophilization.

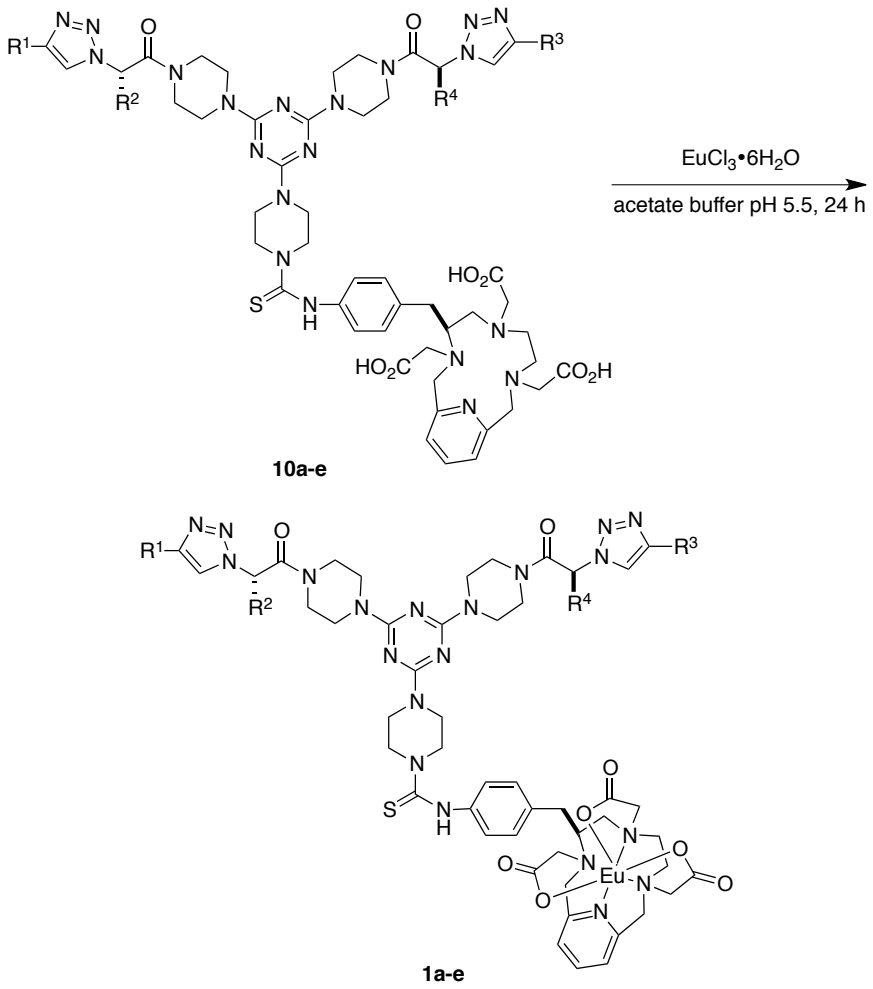


Table S1. Characterization of Compounds **5**.

com'd	sequence (R^1R^2-R^3R^4)	retention time (min)	chemical formula	$[M+H]^+$ calculated	$[M+H]^+$ found
1a	RI-RI	12.2	$C_{64}H_{93}EuN_{26}O_8S$	1539.7	1539.4
1b	SY-KT	10.1	$C_{63}H_{82}EuN_{21}O_{11}S$	1494.6	1494.4
1c	KG-SG	9.3	$C_{54}H_{72}EuN_{21}O_9S$	1344.5	1344.4
1d	TG-KI	11.0	$C_{59}H_{82}EuN_{21}O_9S$	1414.6	1414.4
1e	2mor	10.9	$C_{40}H_{51}EuN_{12}O_8S$	1013.3	1014.0

APPENDIX E

GENERAL EXPERIMENTAL PROCEDURES

E. Molecular Modeling

Templates For Cα – Cβ Distances And Overlay

Templates for ideal type I β-turns, and for γ-turns were obtained from standard torsion angles. A standard template for overlays with an α-helix was obtained from Discovery Studio 2.5, and a β-sheet template for overlays was obtained by β-sheet builder.

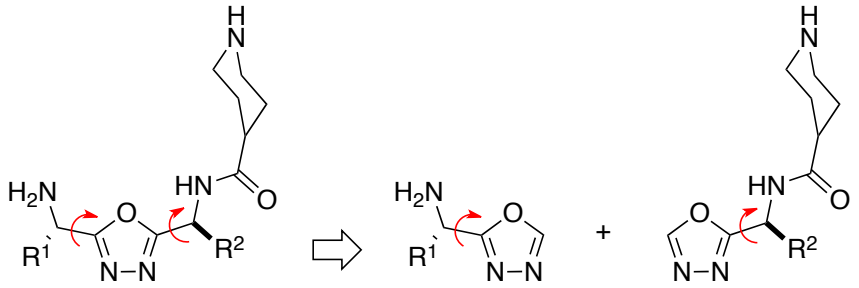
Procedure For Overlays

The contracted and extended *Cβ - Cβ* distances of compound **2** were obtained by rotation of the bonds (red arrows shown below). The possible secondary structures mimicked by compound **2** were obtained by comparing the values between the *Cβ - Cβ*distances for compound **2** against the distances in common secondary structures shown in table 1. After identifying possible secondary structures that compound **2** can mimic, the model of compound **2** was overlaid with the model of the secondary structure. Free rotation about the bonds was allowed to give a good matching of the *Cβ* atoms on compound **1** with the secondary structure. The overlays for compounds **3**, **4, 5,** and **6** were also obtained in a similar way.

Procedures For Calculating Energy Barriers

Reaction path calculations were performed at the B3LYP level of theory with the 6-31G(d') basis set. Full geometry optimizations were performed for each fragment and stationary points were verified by frequency calculations and water solvation calculations. All B3LYP calculations were performed using Gaussian 03.

The energy barriers for rotation of the bonds (red arrows) in compound **2** were calculated on fragments **2A** and **2B**. The energy barrier for the whole structure was calculated by combination of the results for the fragments. The energy barriers for compounds **5** and **6** were also obtained in a similar way. For compounds **3** and **4**, the whole structures were used for the calculation.



2 **2A** **2A**

Procedure For Quenched Molecular Dynamics Studies

NAMD was used for the molecular simulations performed in this work (compounds **2**, **5** and **6**). Explicit atom representations were used throughout the study. The protein structure files (PSF) for all the peptidomimetics were built using Discovery Studio 2.5 (Accelrys Inc) using the CHARMm force field.

Quenched molecular dynamics simulations were performed using the CHARMm force field as implemented in Discovery Studio 2.5. All four molecules were modeled as neutral compounds in a dielectric continuum of 80 (simulating H_2O). Thus, the starting conformers were minimized using 3000 steps of conjugate gradient. The minimized structures were then subjected to heating, equilibration, and dynamics simulation. Throughout, the equations of motions were integrated using the Verlet algorithm with a time step 1 fs. Each peptidomimetic was heated to 1000 K over 10 ps and equilibrated for another 10 ps at 1000 K, then molecular dynamics runs were performed for a total time of 600 ps with trajectories saved every 1 ps. The resulting 600 structures were thoroughly minimized using 1000 steps of SD followed by 3000 steps of conjugate

gradient. Structures with energies less than 0.3 (compound **2** and **5**) and 1.0 (compound **6**) kcal mol^{-1} relative to the global minimum were selected for further analysis.

The VMD package was again used to display, overlay, and classify the selected structures into conformational groups. The best clustering was obtained using a grouping method based on calculation of RMS deviation of a subset of atoms, in this study these were the $C\alpha$- and $C\beta$- atoms. Thus, threshold cutoff values 0.3 Å were selected to obtain families with reasonable homogeneity. The lowest energy conformation from each family was considered to be a typical representative of the family as a whole.

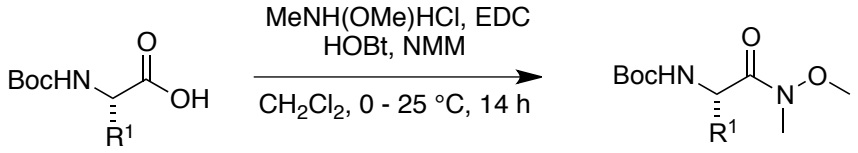
F. Syntheses for 1,3-Butadyne-based Peptidomimetics

General Procedure for Compound 7a - h

To a solution of Boc-L-amino acid (1.0 equiv) in dichloromethane (0.24 M) were added HOBt (1.1 equiv) and EDC (1.2 equiv) at 0 ^{o}C. The solution was stirred at 0 ^{o}C for 15 min and then *N,O*-dimethylhydroxylamine hydrochloride salt (1.15 equiv) and *N*-methyl morpholine (1.2 equiv) were added. The reaction mixture was stirred at 25 ^{o}C for 14 h. After the solvent was removed under vacuum, the resulting residue was partitioned between 1*N* HCl (aq.) and EtOAc. The phases were separated and the organic layer was washed with 1*N* HCl (aq.), followed by saturated $NaHCO_3$ and brine, and then dried over $MgSO_4$. After completely removing the solvent, the Weinreb's amides **7a - d, f - h** were purified by flash chromatography.

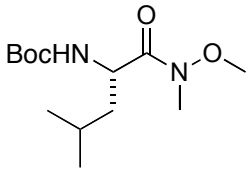
For compound **7e**, the obtained Boc-Tyr-Weinreb amide was used for next step without further purification. To a solution of Boc-Tyr-Weinreb amide (1.0 equiv) in DMF (0.25 M) was added imidazole (3.0 equiv) and TBDPS-Cl (1.7 equiv) at 0 ^{o}C. The reaction mixture was stirred at 25 ^{o}C for 18 h. The reaction mixture was diluted with H_2O and extracted with EtOAc. The combined organic phases were dried over $MgSO_4$. After completely removing the solvent, the **7e** was purified by flash chromatography.

Scheme S3. Synthesis of Compounds **7a - h**.



7a - h

Preparation of Compound 7a



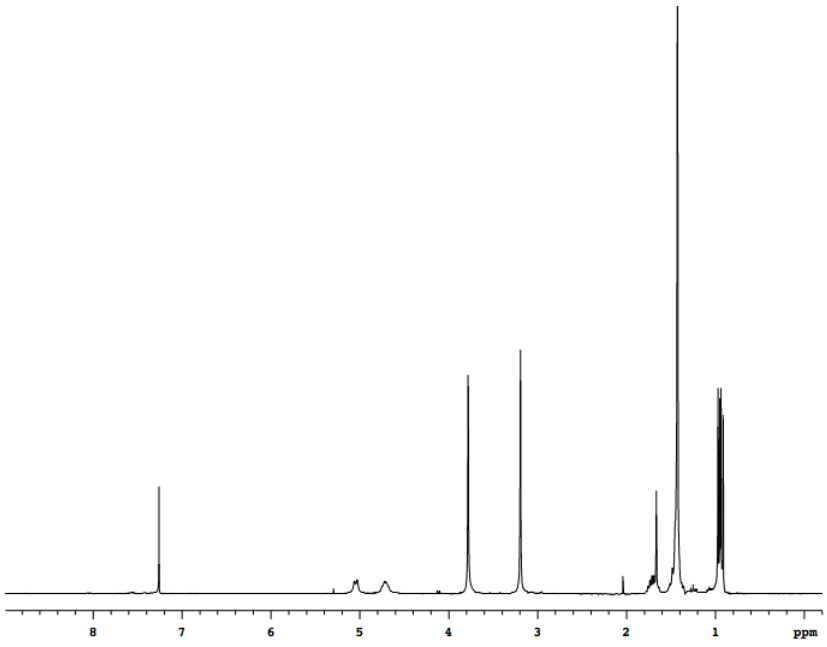
7a

Compound **7a** was prepared from Boc-Leu-OH (9.00 g, 38.9 mmol). Flash chromatography (1:4 EtOAc/Hexanes) afforded 10.7 g (quant.) **7a** as a white solid.

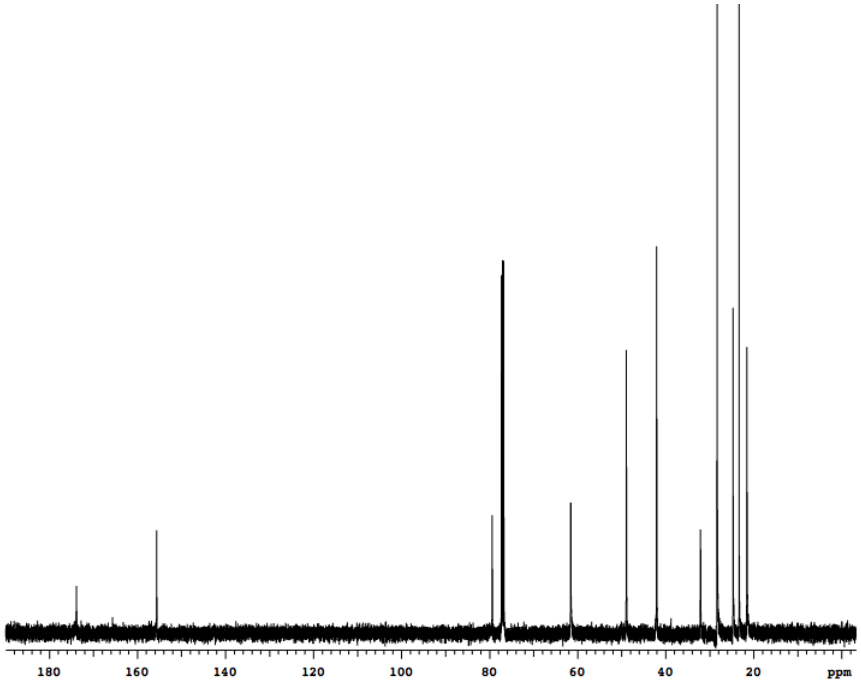
^{1}H NMR (500 MHz, $CDCl_3$) δ 5.04 (d, 1H, J = 14.0 Hz), 4.72 (br, 1H), 3.78 (s, 3H), 3.19 (s, 3H), 1.76-1.66 (m, 2H), 1.42 (br, 10H), 0.96 (d, J = 11.0 Hz), 0.92 (d, J = 11.0 Hz)

^{13}C NMR (125 MHz, $CDCl_3$) δ 173.85, 155.61, 79.40, 61.54, 48.92, 42.04, 32.08, 28.32, 24.69, 23.32, 21.54

MS (ESI, m/z) calcd for $C_{13}H_{27}N_2O_4$ $(M+H)^+$ 275.20, found 275.20

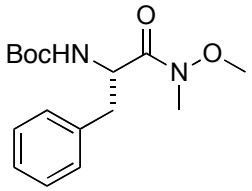


^{1}H NMR of **7a**



^{13}C NMR of **7a**

Preparation of Compound 7b



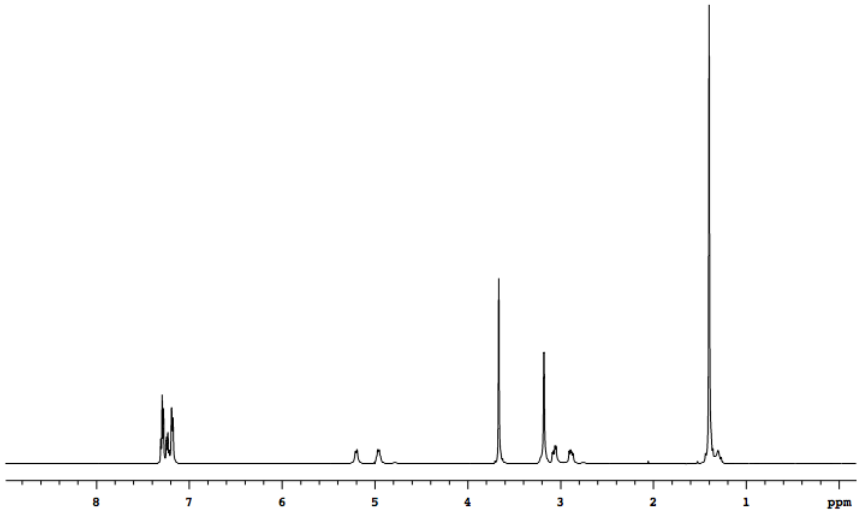
7b

Compound **7b** was prepared from Boc-Phe-OH (4.00 g, 15.0 mmol). Flash chromatography (1:5 EtOAc/Hexanes) afforded 4.65 g (quant.) **7b** as a yellowish oil.

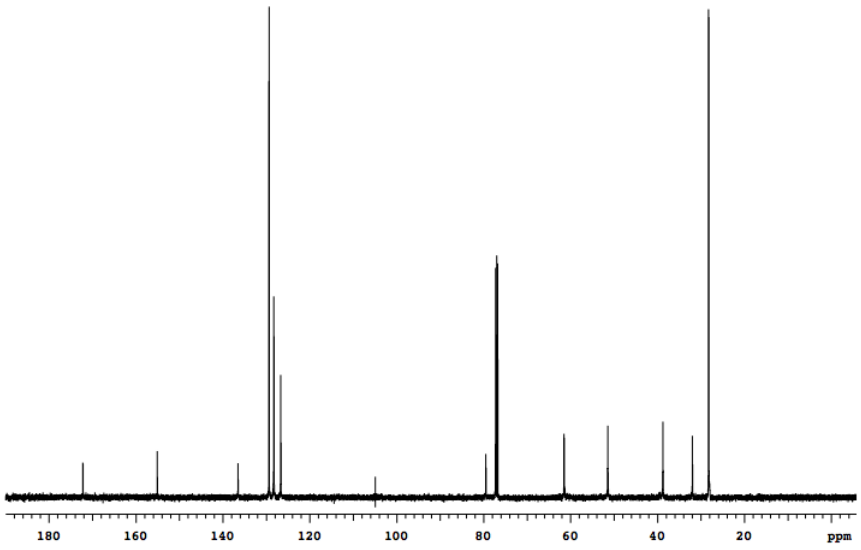
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.27 (t, 2H, J = 7.5 Hz), 7.20 (t, 1H, J = 7.5 Hz), 7.26 (d, 2H, J = 7.0), 5.17 (d, 1H, J = 7.0 Hz), 4.94 (br, 1H), 3.64 (s, 3H), 3.15 (s, 3H), 3.04 (dd, 1H, J = 4.5, 13.5 Hz), 2.86 (dd, 1H, J = 7.0, 13.0 Hz), 1.37 (s, 9H)

^{13}C NMR (125 MHz, $CDCl_3$) δ 172.18, 155.07, 136.50, 129.37, 128.25, 126.66, 79.47, 61.48, 51.42, 38.74, 31.96, 28.23 (107.25 is noise from NMR.)

MS (ESI, m/z) calcd for $C_{16}H_{25}N_2O_4$ $(M+H)^+$ 309.18, found 309.17

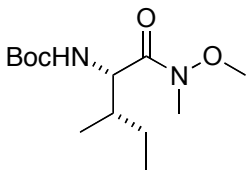


^{1}H NMR of **7b**



^{13}C NMR of **7b**

Preparation of Compound 7c



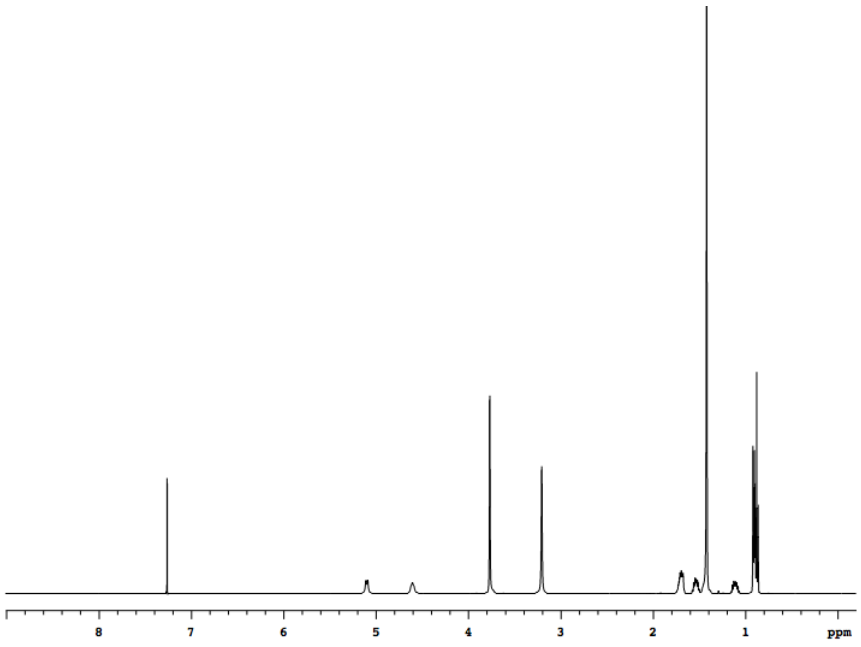
7c

Compound **7c** was prepared from Boc-Ile-OH (5.00 g, 21.6 mmol). Flash chromatography (1:3 EtOAc/Hexanes) afforded 5.93 g (quant.) **7c** as a colorless oil.

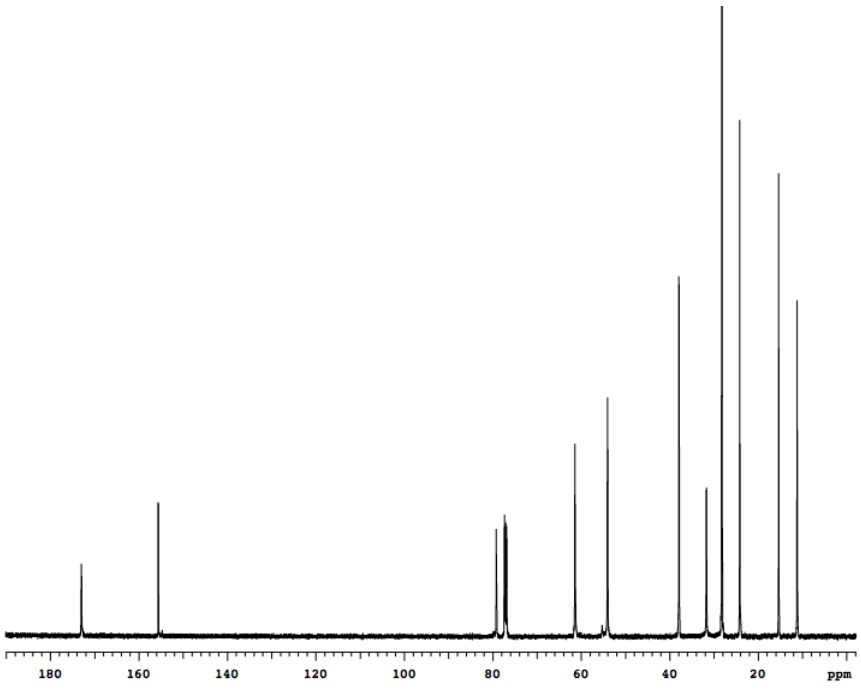
^{1}H NMR (500 MHz, $CDCl_3$) δ 5.11 (d, 1H, J = 10.0 Hz), 4.61 (br, 1H), 3.78 (s, 3H), 3.22 (s, 3H), 1.75-1.69 (m, 2H), 1.59-1.51 (m, 1H), 1.43 (s, 9H), 0.92 (d, 3H, J = 6.5 Hz), 0.89 (t, 3H, J = 8.5 Hz)

^{13}C NMR (125 MHz, $CDCl_3$) δ 172.89, 155.52, 79.12, 61.32, 53.94, 37.80, 31.60, 28.13, 24.07, 15.27, 11.12

MS (ESI, m/z) calcd for $C_{13}H_{27}N_2O_4$ $(M+H)^+$ 275.20, found 275.19

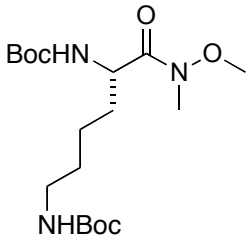


^{1}H NMR of **7c**



^{13}C NMR of **7c**

Preparation of Compound 7d



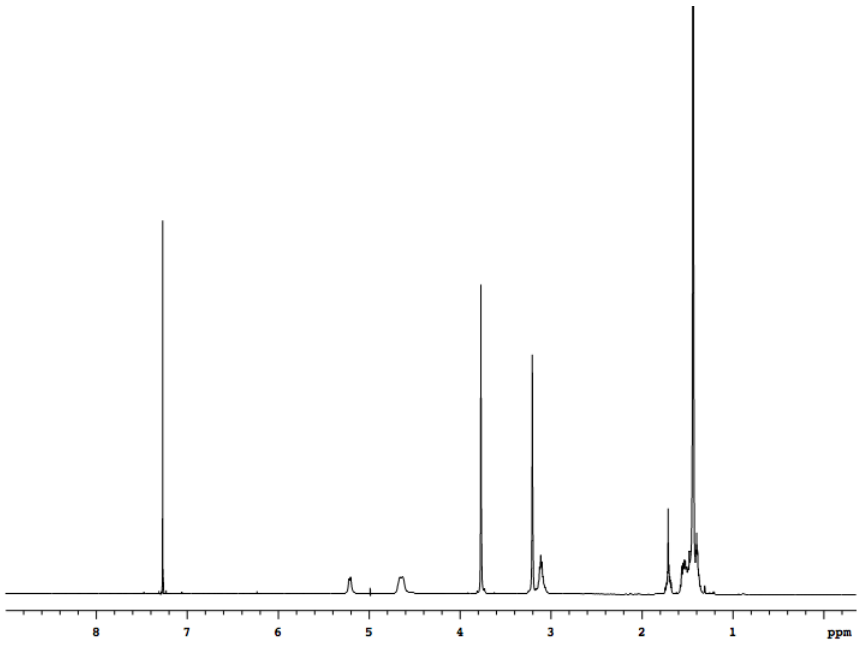
7d

Compound **7d** was prepared from Boc-Lys(Boc)-OH (5.00 g, 14.4 mmol). Flash chromatography (1:3 EtOAc/Hexanes) afforded 5.40 g (96 %) **7d** as a colorless oil.

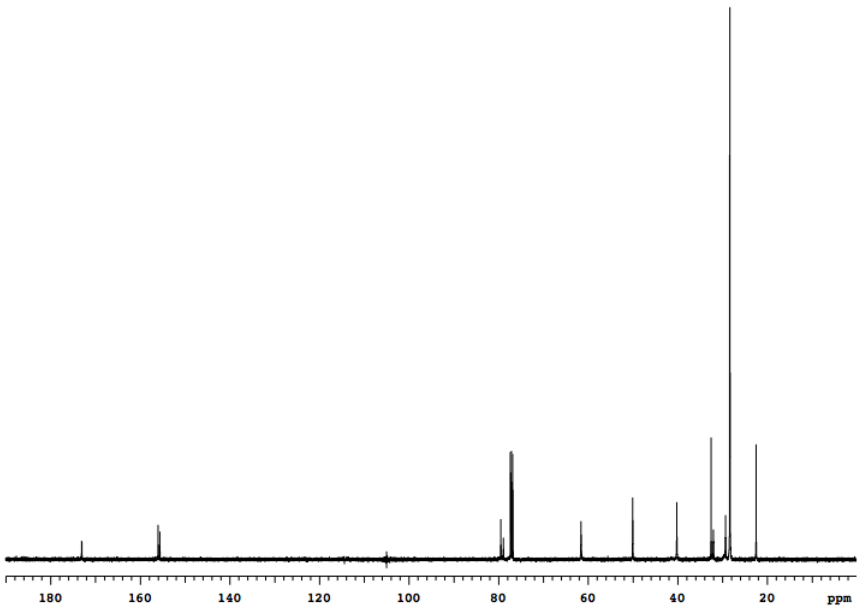
^{1}H NMR (500 MHz, $CDCl_3$) δ 5.21 (d, 1H, J = 8.5 Hz), 4.66 (br, 1H), 4.63 (br, 1H), 3.77 (s, 3H), 3.20 (s, 3H), 3.15-3.09 (m, 2H), 1.75-1.68 (m, 2H), 1.58-1.49 (m, 2H), 1.44 (s, 18H), 1.42-1.37 (m, 2H)

^{13}C NMR (125 MHz, $CDCl_3$) δ 172.97, 155.91, 155.56, 79.42, 78.83, 61.49, 49.96, 40.12, 32.46, 31.94, 29.24, 28.31, 28.25, 22.42 (107.25 is noise from NMR.)

MS (ESI, m/z) calcd for $C_{18}H_{36}N_3O_6$ $(M+H)^+$ 390.26, found 390.27

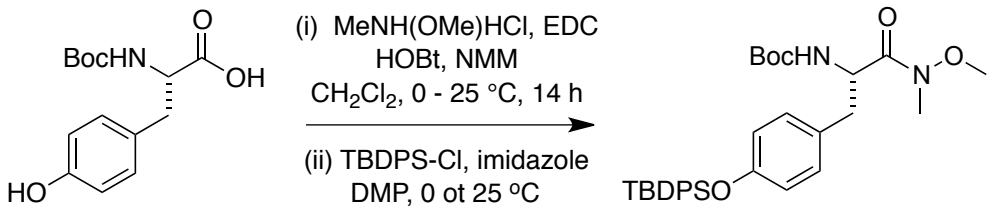


^{1}H NMR of **7d**



^{13}C NMR of **7d**

Preparation of Compound 7e



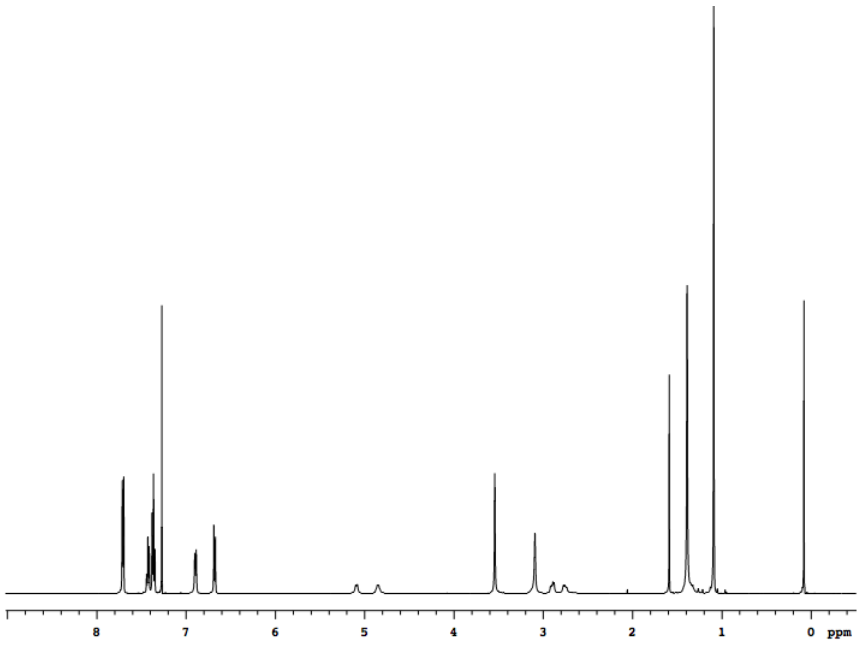
7e

Compound **7e** was prepared from Boc-Tyr-OH (3.1 g, 11.0 mmol). Flash chromatography (1:6 EtOAc/Hexanes) afforded 2.5 g (71 % over 2 steps) **7e** as a white solid.

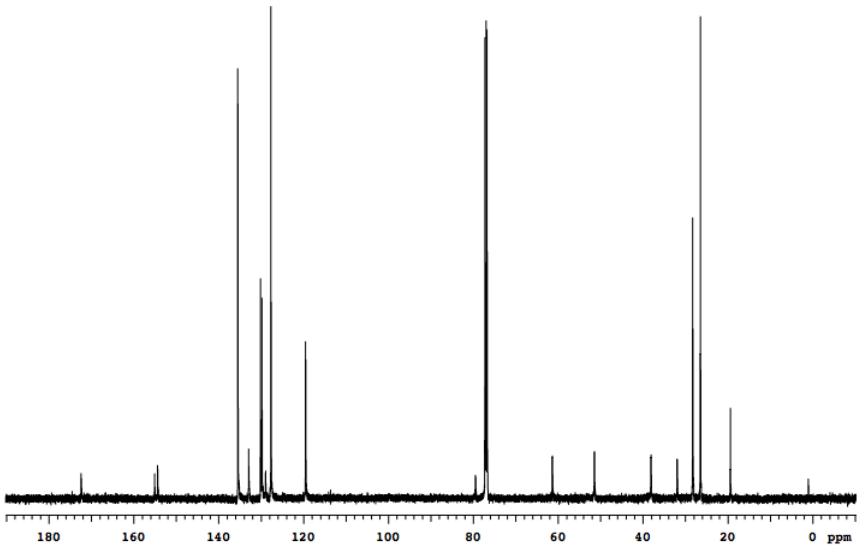
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.71 (d, 4H, J = 7.0 Hz), 7.43 (dd, 2H, J = 7.5, 7.5 Hz), 7.36 (dd, 4H, J = 7.5, 7.5 Hz), 6.89 (d, 2H, J = 8.0 Hz), 6.68 (d, 2H, 8.5 Hz), 5.09 (d, 1H, J = 8.5 Hz), 4.86-4.84 (m, 1H), 3.54 (s, 3H), 3.09 (s, 3H), 2.90 (dd, 1H, J = 6.5, 13.5 Hz), 1.39 (s, 9H), 1.09 (s, 9H)

^{13}C NMR (125 MHz, $CDCl_3$) δ 172.46, 155.15, 154.44, 135.53, 132.98, 130.20, 129.86, 128.98, 127.74, 119.56, 79.52, 61.42, 51.49, 38.16, 31.98, 28.34, 26.53, 19.47

MS (ESI, m/z) calcd for $C_{32}H_{42}N_2NaO_5Si$ $(M+Na)^+$ 585.28, found 585.26

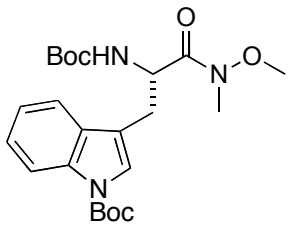


^{1}H NMR of **7e**



^{13}C NMR of **7e**

Preparation of Compound 7f



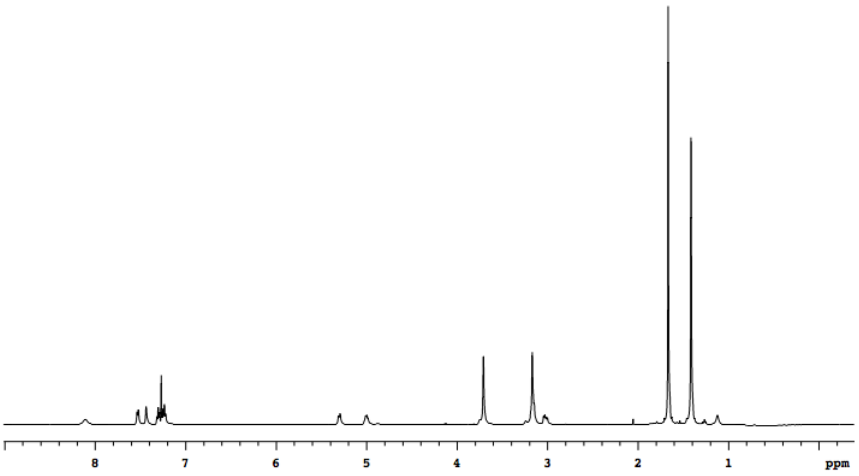
7f

Compound **7f** was prepared from Boc-Trp(Boc)-OH (3.00 g, 9.9 mmol). Flash chromatography (1:7 EtOAc/Hexanes) afforded 3.30 g (quant.) **7f** as a white solid.

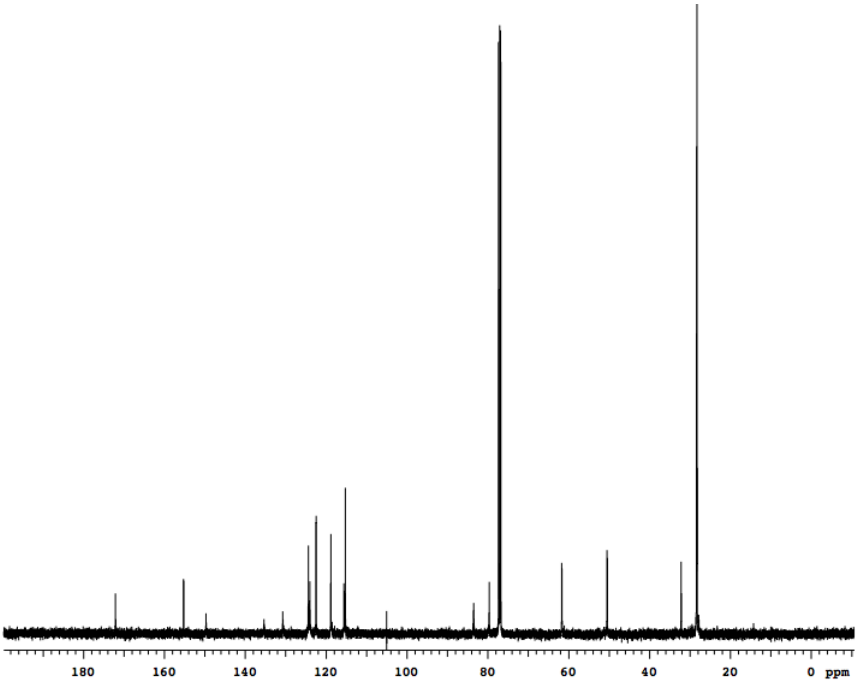
^{1}H NMR (500 MHz, $CDCl_3$) δ 8.11 (br, 1H), 7.53 (d, 1H, J = 7.5 Hz), 7.44 (s, 1H), 7.30 (t, 1H, J = 7.5 Hz), 7.23 (t, 1H, 8.0 Hz), 5.30 (d, 1H, J = 7.5 Hz), 5.01-4.96 (m, 1H), 3.71 (s, 3H), 3.17 (s 3H), 3.17-3.15 (m, 1H), 3.02 (dd, 1H, J = 6.5, 14.0 Hz), 1.67 (s, 9H), 1.41 (s, 9H)

^{13}C NMR (125 MHz, $CDCl_3$) δ 172.03, 155.16, 149.62, 135.29, 130.65, 124.30, 123.95, 122.41, 118.76, 115.49, 115.19, 83.42, 79.58, 61.62, 50.42, 32.10, 28.29, 28.18, 27.71 (107.25 is noise from NMR.)

MS (ESI, m/z) calcd for $C_{23}H_{34}N_3O_6$ $(M+H)^+$ 448.24, found 448.25

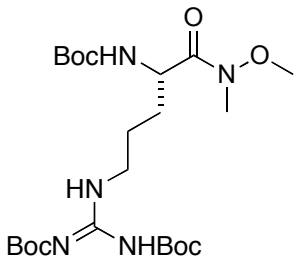


^{1}H NMR of **7f**



^{13}C NMR of **7f**

Preparation of Compound 7g



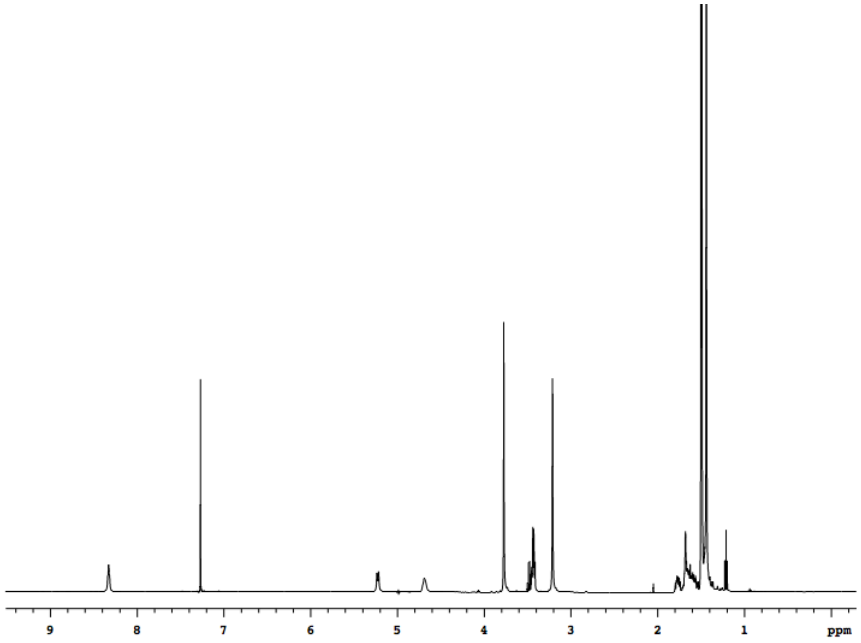
7g

Boc-Arg(Boc)$_2$-OH was prepared from H-Orn-OH· HCl according to the literature procedure. Compound **7g** was prepared from Boc-Arg(Boc)$_2$-OH (3.10 g, 6.5 mmol). Flash chromatography (1:3 to 1:2 EtOAc/Hexanes) afforded 3.38 g (68 %) **7g** as a white solid.

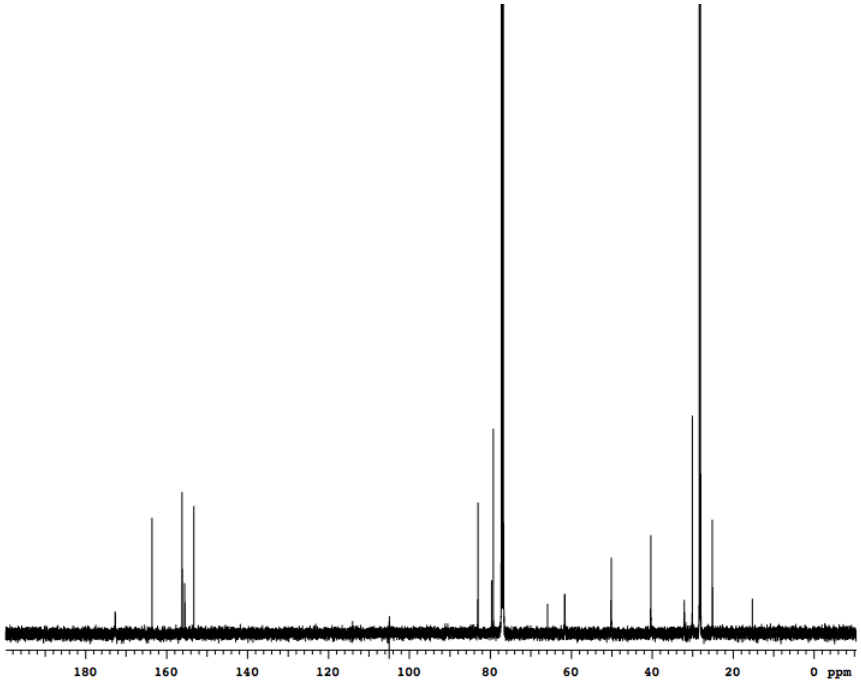
^{1}H NMR (500 MHz, $CDCl_3$) δ 8.33 (s, 1H), 5.23 (d, 1H, J = 9.0 Hz), 4.69 (br, 1H), 3.77 (s, 3H), 3.43 (td, 2H, J = 3.5, 10.0 Hz), 3.21 (s, 3H), 1.80-1.74 (m, 1H), 1.69-1.52 (m, 3H), 1.52 (s, 9H), 1.51 (s, 9H), 1.45 (s, 9H)

^{13}C NMR (125 MHz, $CDCl_3$) δ 172.69, 163.58, 156.14, 155.50, 153.23, 83.05, 79.64, 79.22, 61.60, 50.11, 40.33, 32.04, 30.07, 28.33, 28.27, 28.04, 25.12 (107.25 is noise from NMR.)

MS (ESI, m/z) calcd for $C_{23}H_{44}N_5O_8$ $(M+H)^+$ 518.32, found 518.34

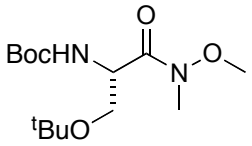


^{1}H NMR of **7g**



^{13}C NMR of **7g**

Preparation of Compound 7h



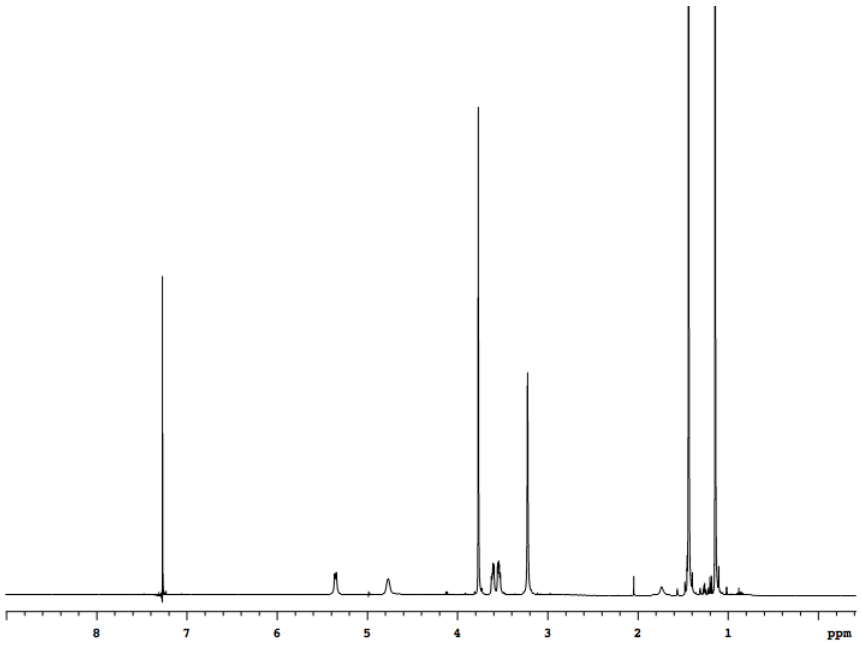
7h

Compound **7h** was prepared from Boc-Ser(^{t}Bu)-OH (3.50 g, 13.4 mmol). Flash chromatography (1:9 EtOAc/Hexanes) afforded 4.06 g (quant.) **7h** as a colorless oil.

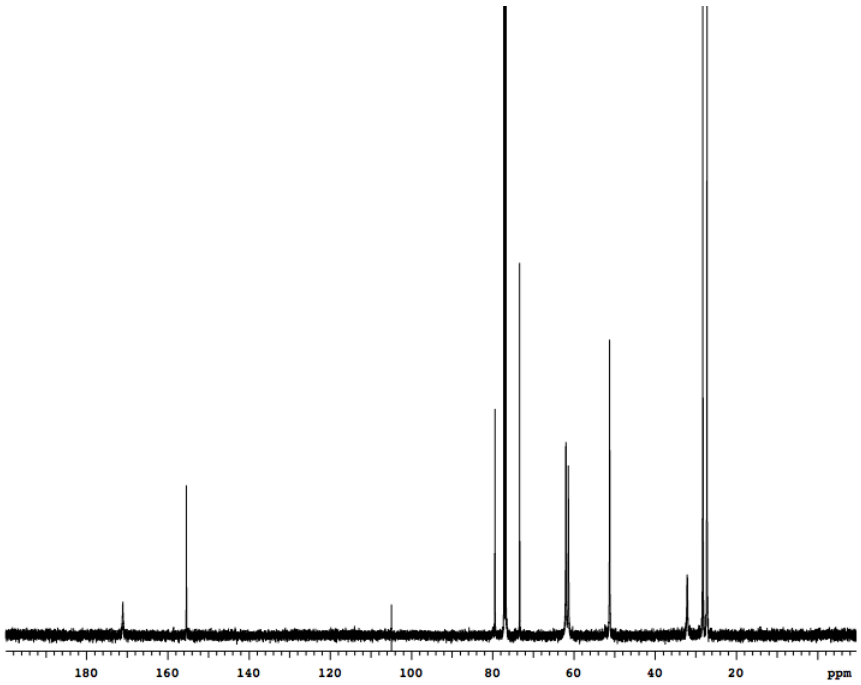
^{1}H NMR (500 MHz, $CDCl_3$) δ 5.35 (d, 1H, J = 9.0 Hz), 4.77 (br, 1H), 3.77 (s, 3H), 3.61 (dd, 1H, J = 5.0, 9.0 Hz), 3.55 (dd, 1H, J = 5.0, 8.5 Hz), 3.22 (s, 3H), 1.44 (s, 9H), 1.14 (s, 9H)

^{13}C NMR (125 MHz, $CDCl_3$) δ 171.06, 155.41, 79.43, 73.36, 61.97, 61.35, 51.19, 32.08, 28.26, 27.23 (107.25 is noise from NMR.)

MS (ESI, m/z) calcd for $C_{14}H_{29}N_2O_5$ $(M+H)^+$ 305.21, found 305.20



^{1}H NMR of **7h**



^{13}C NMR of **7h**

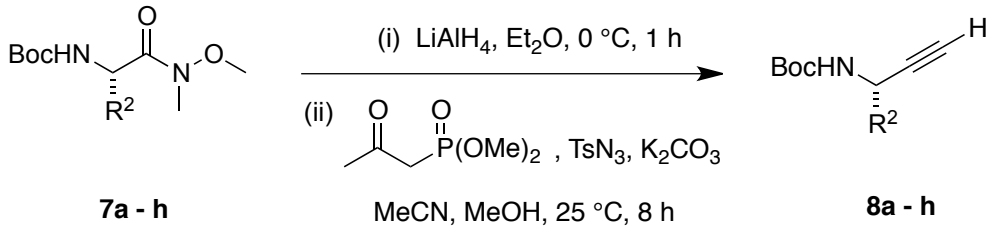
General Procedure for Compounds 8a - h

The alkynes **8a - h** were prepared by a procedure previously described in detail. The Weinreb's amide (1.0 equiv) was dissolved in dry ether (0.15 M) and cooled to 0 °C under N_2 (g). A solution of LiAlH (1.1 equiv, 1.0 M in Et_2O) was added via syringe over a period of 20 min. The resulting reaction mixture was stirred at 0 °C for 1 h and then quenched by dropwise addition of 5% $KHSO_4$ (aq.). The phases were separated and the aqueous phase was extracted with ether. The combined organic phases were dried over $MgSO_4$, and then the solvent was removed to get the amino aldehyde, which was used immediately for next step without further purification.

To a solution of TsN_3 (1.3 equiv) in dry MeCN (0.1 M) were added K_2CO_3 (3.0 equiv) and dimethyl 2-oxopropylphosphonate (1.3 equiv), and was stirred at 25 °C for 2 hours. The amino acid aldehyde obtained above was dissolved in MeOH (0.37 M) and this solution was added into the reaction mixture all in once. The resulting solution was

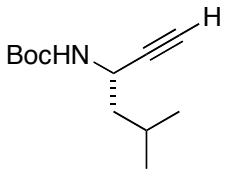
stirred at 25 °C for 18 hours before the solvent was removed. The residue was partitioned between ether and water. The layers were separated and the aqueous phase was extracted with ether. The combined organic phases were dried over $MgSO_4$. After completely removing the solvent, the amino alkynes **8a - h** were purified by flash chromatography.

Scheme S4. Synthesis of compounds **8a - h**.



7a - h **8a - h**

Preparation of Compound 8a



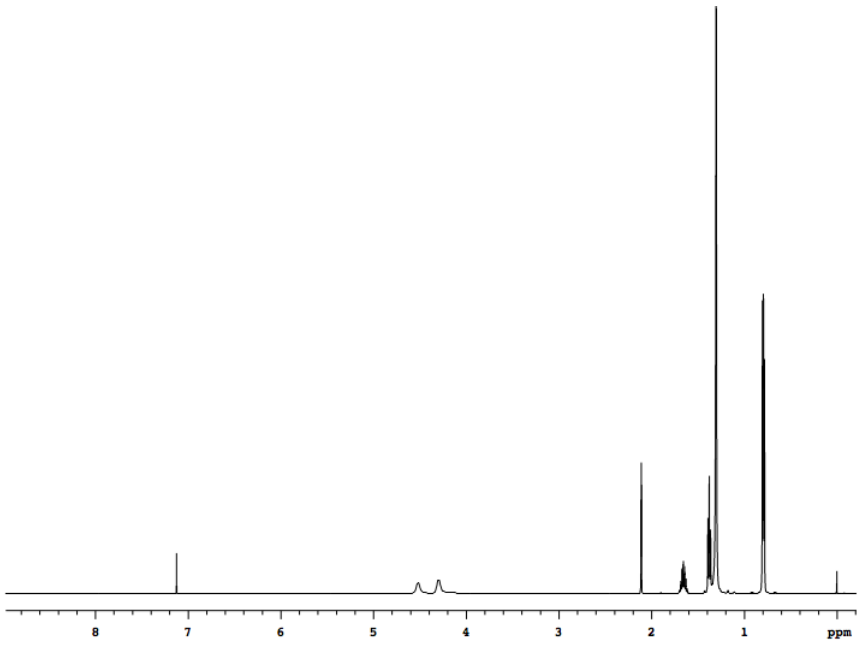
8a

Compound **8a** was prepared from **7a** amide (3.90 g, 18.1 mmol). Flash chromatography (1:6 EtOAc/Hexanes) afforded 2.57 g (67 %) **8a** as a white solid.

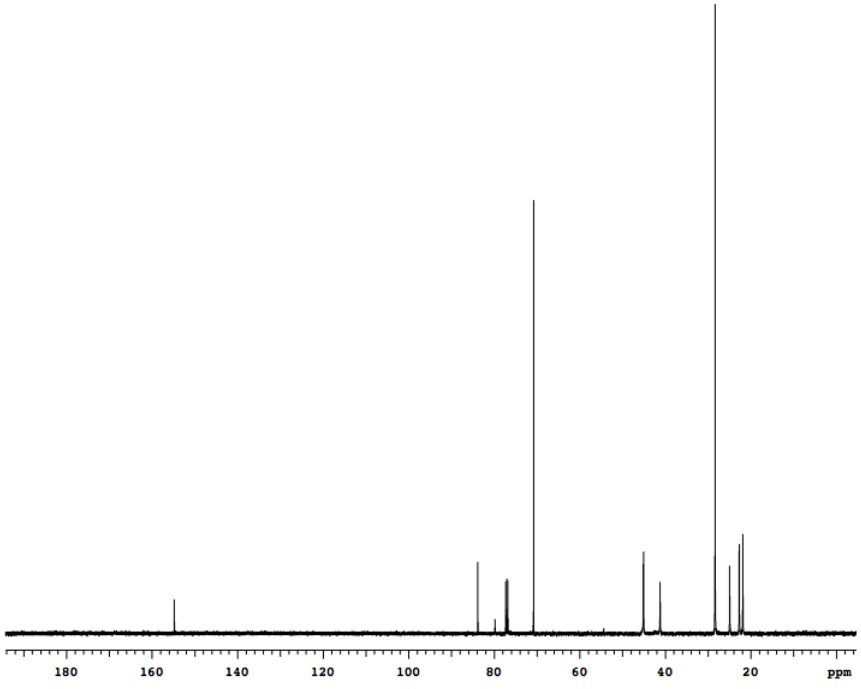
^{1}H NMR (500 MHz, $CDCl_3$) δ 4.52(br, 1H), 4.30 (br, 1H), 2.11 (d, 1H, J = 3.0 Hz), 1.70-1.62 (m, 1H), 1.38 (t, 2H, J = 7.5 Hz), 1.30 (s, 9H), 0.80 (d, 3H, J = 5.0 Hz), 0.79 (d, 3H, J = 5.0 Hz)

^{13}C NMR (500 MHz, $CDCl_3$) δ 154.78, 83.82, 79.79, 70.77, 45.09, 41.20, 28.35, 24.95, 22.70, 21.85

MS (ESI, m/z) calcd for $C_{13}H_{27}N_2O_4$ $(M+H)^+$ 212.17, found 212.16

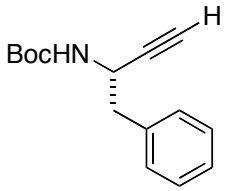


^{1}H NMR of **8a**



^{13}C NMR of **8a**

Preparation of Compound 8b



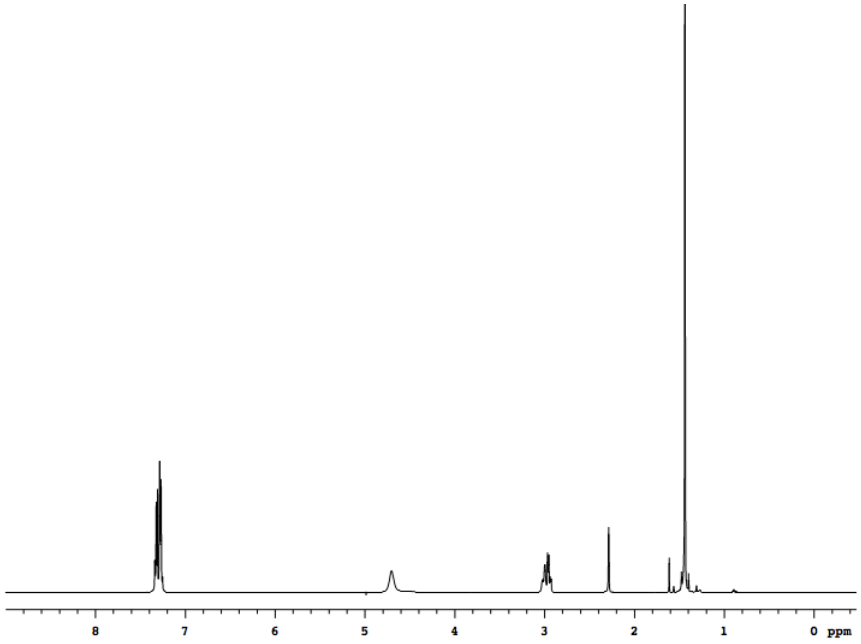
8b

Compound **8b** was prepared from **7b** amide (4.00 g, 13.0 mmol). Flash chromatography (1:5 EtOAc/Hexanes) afforded 2.07 g (65 %) **8b** as a white solid.

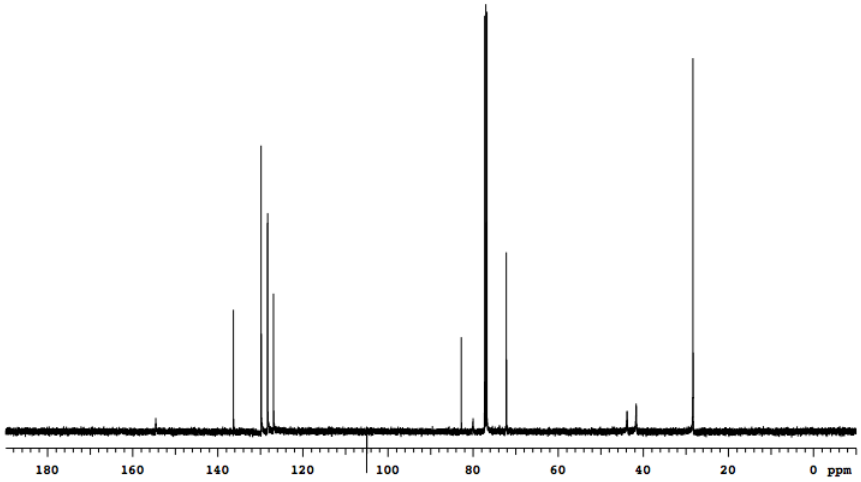
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.34-7.25 (m, 5H), 4.70 (br, 2H), 3.02-3.00 (m, 1H), 2.95 (dd, 1H, J = 7.0, 13.5 Hz), 2.29 (s, 1H), 1.45 (s, 9H)

^{13}C NMR (125 MHz, $CDCl_3$) δ 154.56, 136.29, 129.78, 128.28, 126.88, 82.74, 80.00, 72.15, 43.80, 41.66, 28.30 (107.25 is noise from NMR.)

MS (ESI, m/z) calcd for $C_{15}H_{20}NO_2$ $(M+H)^+$ 246.15, found 246.14

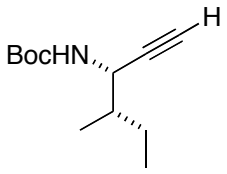


^{1}H NMR of **8b**



^{13}C NMR of **8b**

Preparation of Compound 8c



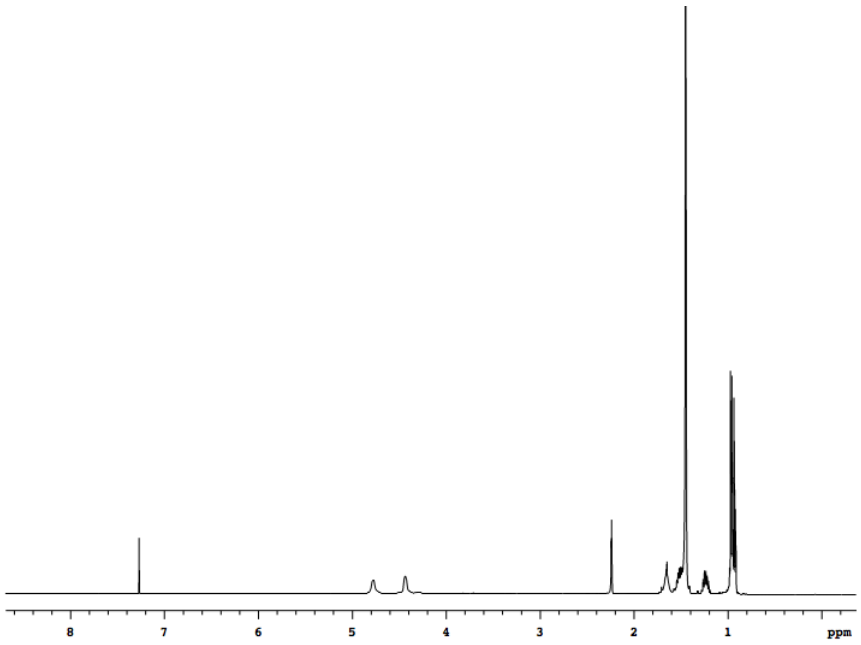
8c

Compound **8c** was prepared from **7c** (5.00 g, 18.2 mmol). Flash chromatography (1:9 EtOAc/Hexanes) afforded 2.41 g (63 %) **8c** as a colorless oil.

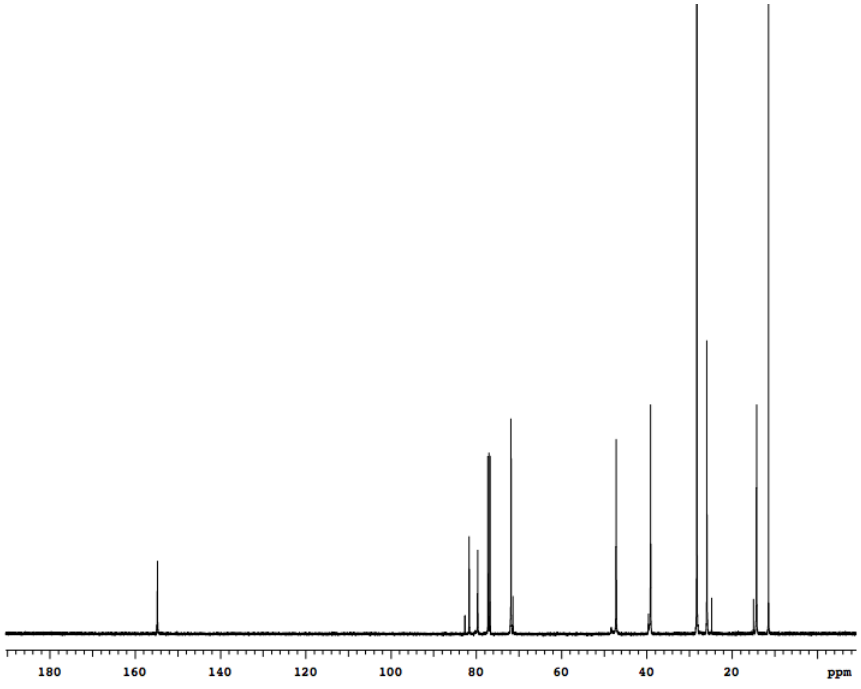
^{1}H NMR (500 MHz, $CDCl_3$) δ 4.77 (br, 1H), 4.44 (br, 1H), 2.24 (s, 1H), 1.65 (br, 1H), 1.55-1.49 (m, 1H), 1.45 (s, 9H) 1.28-1.19 (m, 1H), 0.97 (d, 3H, J = 7.0 Hz), 0.94 (t, 3H, J = 8.5 Hz)

^{13}C NMR (125 MHz, $CDCl_3$) δ 154.84, 81.72, 79.73, 71.91, 47.24, 39.17, 28.33, 25.98, 14.31, 11.51

MS (ESI, m/z) calcd for $C_{12}H_{22}NO_2$ $(M+H)^+$ 212.16, found 212.16

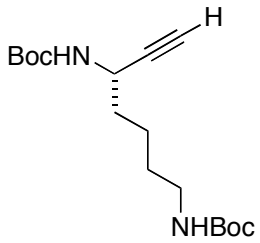


^{1}H NMR of **8c**



^{13}C NMR of **8c**

Preparation of Compound 8d



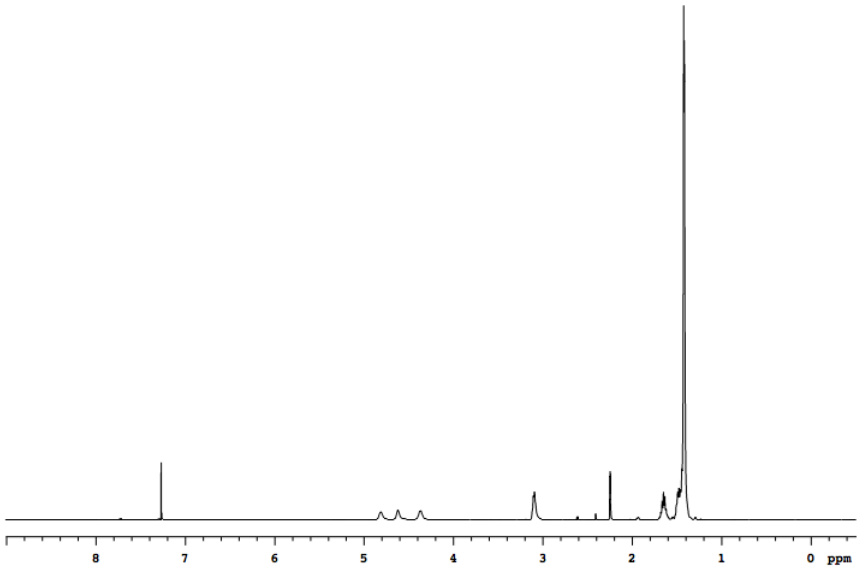
8d

Compound **8d** was prepared from **7d** (5.20 g, 13.4 mmol). Flash chromatography (1:9 EtOAc/Hexanes) afforded 2.18 g (51 %) **8d** as a white solid.

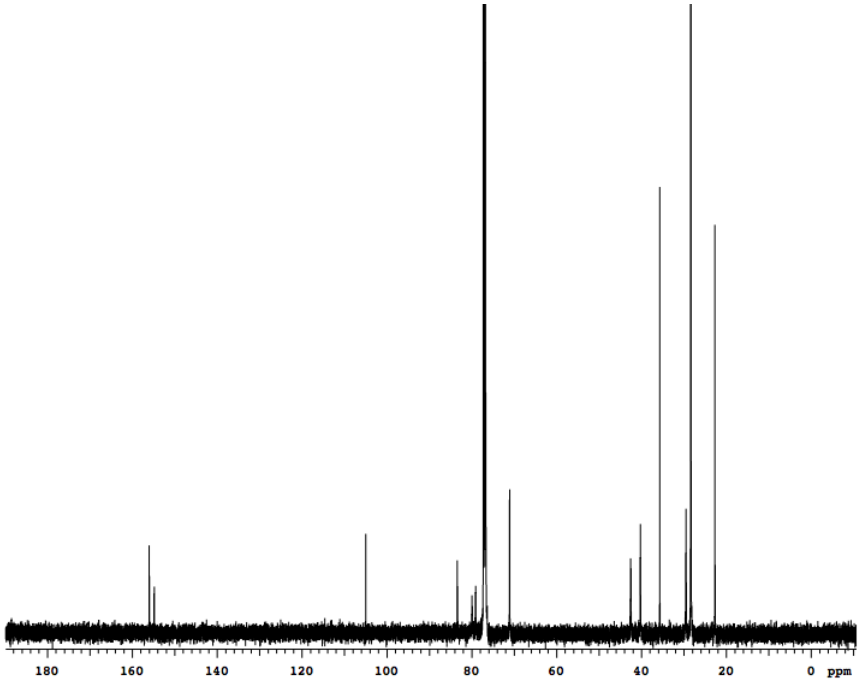
^{1}H NMR (500 MHz, $CDCl_3$) δ 4.74 (br, 1H), 4.56 (br, 1H), 4.39 (br, 1H), 3.14-3.12 (m, 2H), 2.26 (s, 1H), 1.72-1.63 (m, 2H), 1.55-1.45 (m, 4H), 1.45 (s, 9H), 1.44 (s, 9H)

^{13}C NMR (125 MHz, $CDCl_3$) δ 155.99, 154.82, 83.39, 79.92, 79.09, 71.07, 42.53, 40.26, 35.70, 29.51, 28.41, 28.34, 22.69 (107.25 is noise from NMR.)

MS (ESI, m/z) calcd for $C_{17}H_{31}N_2O_4$ $(M+H)^+$ 327.23, found 327.20

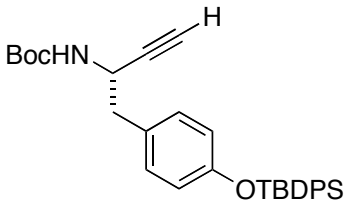


^{1}H NMR of **8d**



^{13}C NMR of **8d**

Preparation of Compound 8e



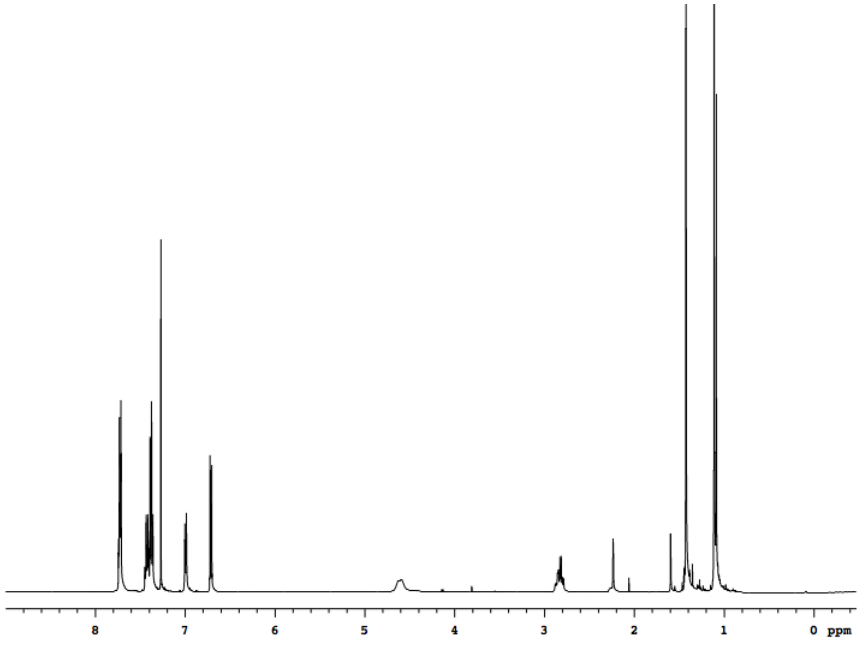
8e

Compound **8e** was prepared from **7e** (2.5 g, 4.4 mmol). Flash chromatography (3:97 EtOAc/Hexanes) afforded 1.50 g (67 %) **8e** as a yellowish oil.

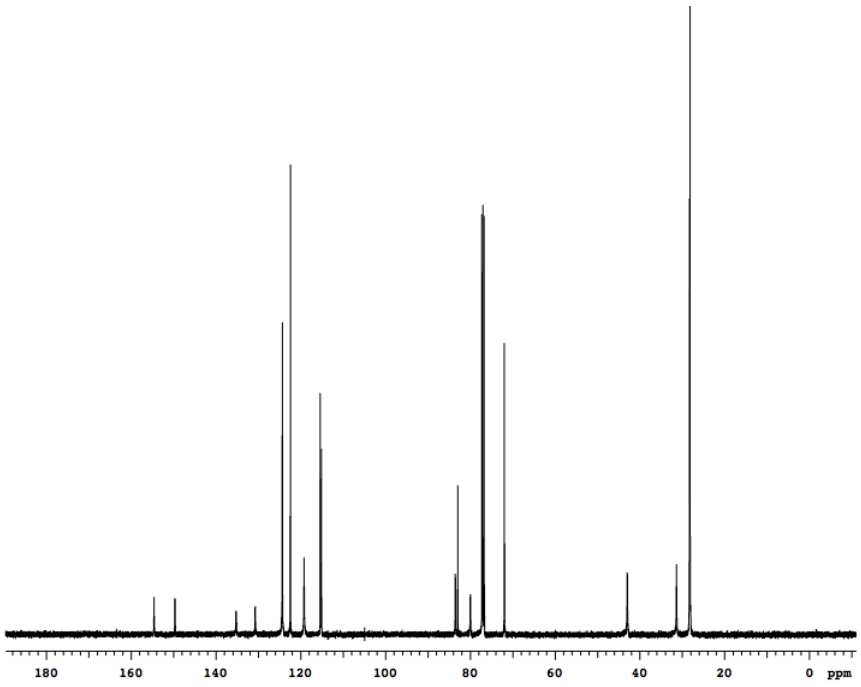
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.72 (d, 4H, J = 8.0 Hz), 7.43 (dd, 2H, J = 7.0, 15.0 Hz), 7.39 (dd, 4H, J = 8.5, 8.5 Hz), 6.99 (d, 2H, J = 8.5 Hz), 6.71 (d, 2H, 8.5 Hz), 4.63 (br, 1H), 4.59 (br, 1H), 2.86 (dd, 1H, J = 5.0, 14.0 Hz), 2.81 (dd, 1H J = 7.0, 13.5 Hz), 2.23 (s, 1H), 1.42 (s, 9H), 1.11 (s, 9H)

^{13}C NMR (125 MHz, $CDCl_3$) δ 154.50, 154.45, 135.42, 135.41, 134.75, 132.86, 130.43, 129.77, 129.38, 128.70, 127.65, 127.53, 119.41, 82.86, 79.74, 71.93, 43.73, 40.69, 28.23, 26.45, 19.35 (107.25 is noise from NMR.)

MS (MALDI, m/z) calcd for $C_{31}H_{37}NNaO_3Si$ $(M+Na)^+$ 522.71, found 522.20

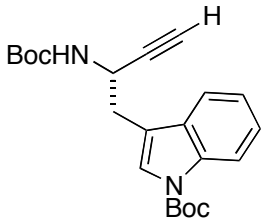


^{1}H NMR of **8e**



^{13}C NMR of **8e**

Preparation of Compound 8f



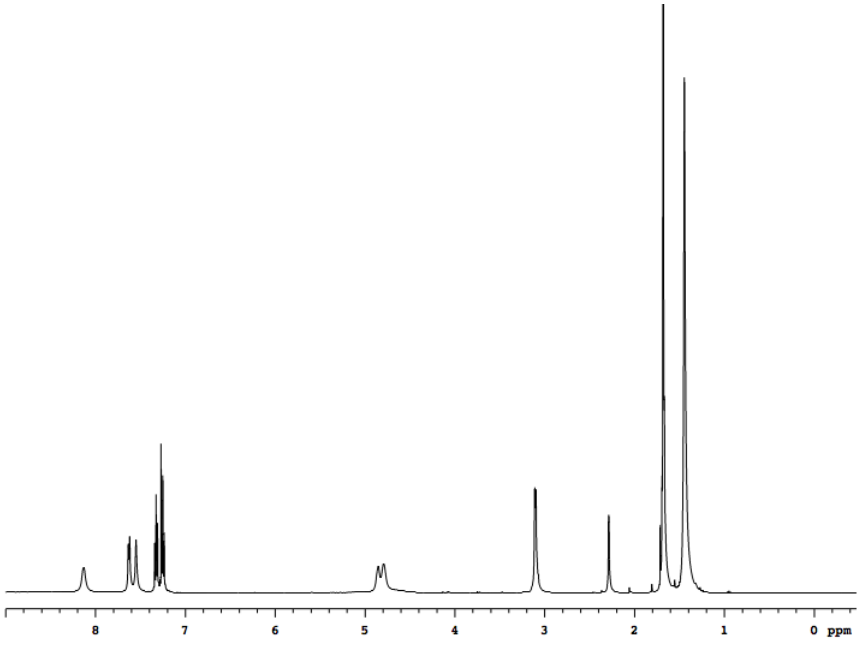
8f

Compound **8f** was prepared from **7f** (3.20 g, 7.2 mmol). Flash chromatography (1:9 EtOAc/Hexanes) afforded 1.90 g (77 %) **8f** as a white solid.

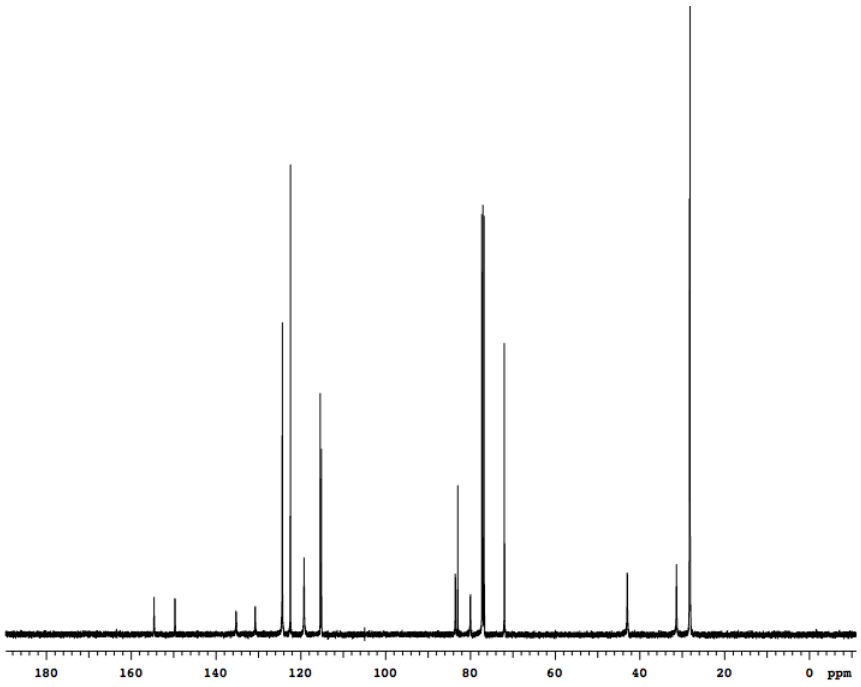
^{1}H NMR (500 MHz, $CDCl_3$) δ 8.13 (br, 1H), 7.63 (d, 1H, J = 6.5 Hz), 7.55 (s, 1H), 7.33 (t, 1H, J = 8.5 Hz), 7.25 (t, 1H, 8.5 Hz), 4.85 (br, 1H), 4.79 (br, 1H), 3.10 (d, 2H, J = 6.5 Hz), 2.29 (s, 1H), 1.68 (s, 9H), 1.44 (s, 9H)

^{13}C NMR (125 MHz, $CDCl_3$) δ 154.60, 149.66, 135.23, 130.73, 124.32, 124.32, 122.43, 119.21, 115.36, 115.13, 83.52, 82.96, 79.96, 71.96, 42.97, 31.35, 28.27, 28.17 (107.25 is noise from NMR.)

MS (ESI, m/z) calcd for $C_{22}H_{29}N_2O_4$ $(M+H)^+$ 385.21, found 385.21

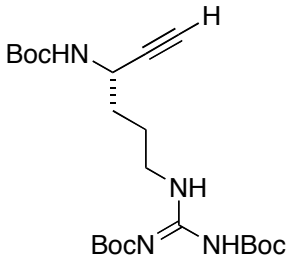


^{1}H NMR of **8f**



^{13}C NMR of **8f**

Preparation of Compound 8g



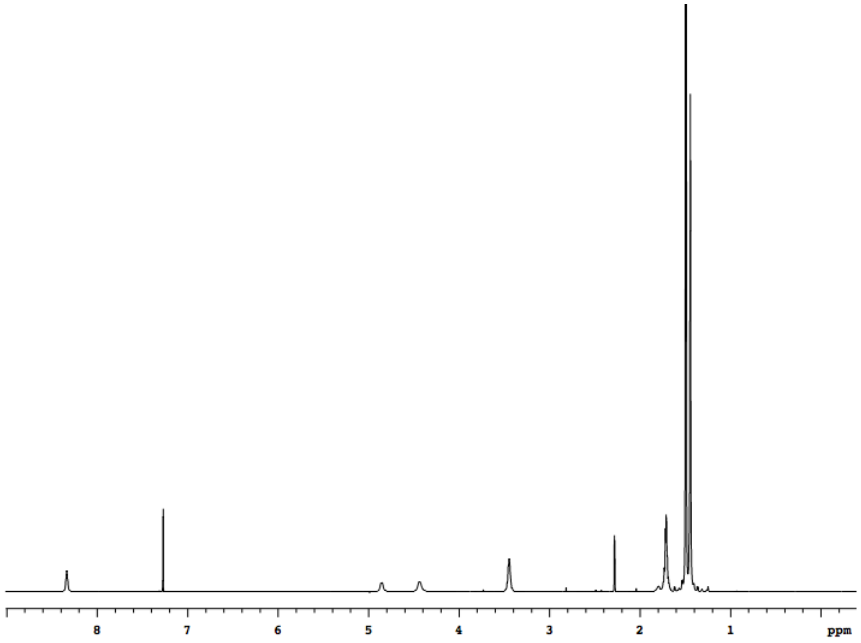
8g

Compound **8g** was prepared from **7g** (1.86 g, 3.6 mmol). Flash chromatography (1:9 EtOAc/Hexanes) afforded 0.50 g (31 %) **8g** as a colorless oil.

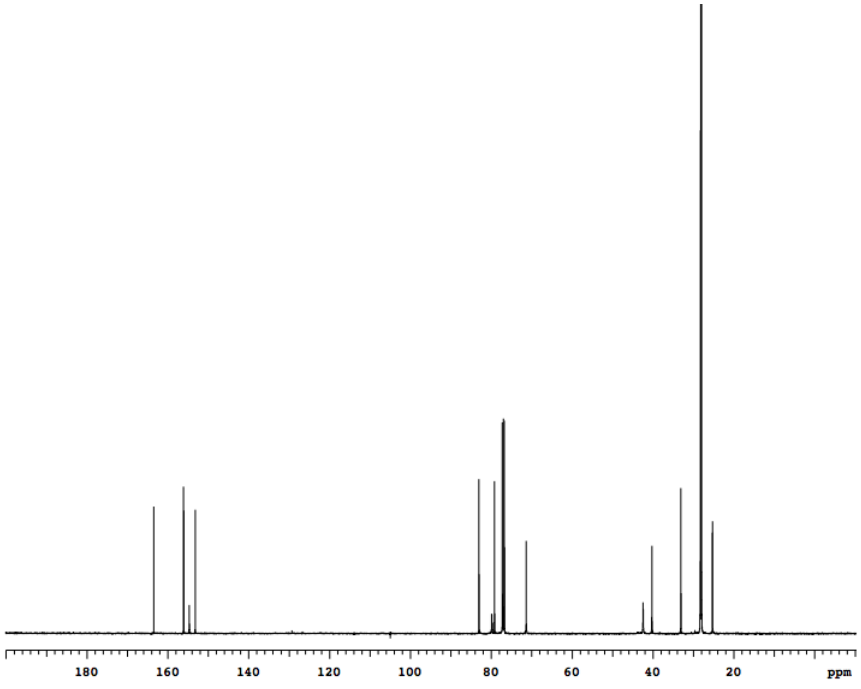
^{1}H NMR (500 MHz, $CDCl_3$) δ 8.34 (s, 1H), 4.86 (d, 1H, J = 6.5 Hz), 4.44 (br, 1H), 3.46-3.44 (m, 2H), 2.28 (s, 1H), 1.73-1.70 (m, 4H), 1.50 (s, 9H), 1.49 (s, 9H), 1.44 (s, 9H)

^{13}C NMR (125 MHz, $CDCl_3$) δ 163.47, 156.08, 154.68, 153.20, 83.03, 79.88, 79.50, 79.18, 71.35, 42.43, 40.24, 33.08, 28.26, 28.21, 27.98, 25.29 (107.25 is noise from NMR.)

MS (MALDI, m/z) calcd for $C_{22}H_{39}N_4O_6$ $(M+H)^+$ 455.29, found 455.23

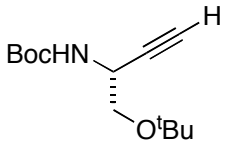


^{1}H NMR of **8g**



^{13}C NMR of **8g**

Preparation of Compound 8h



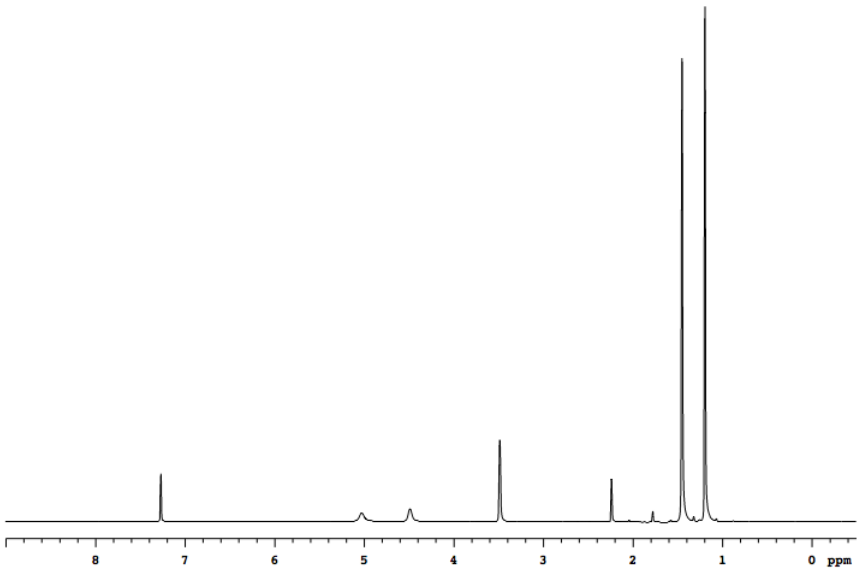
8h

Compound **8h** was prepared from **7h** (2.30 g, 9.2 mmol). Flash chromatography (1:9 EtOAc/Hexanes) afforded 1.50 g (68 %) **8h** as a colorless solid.

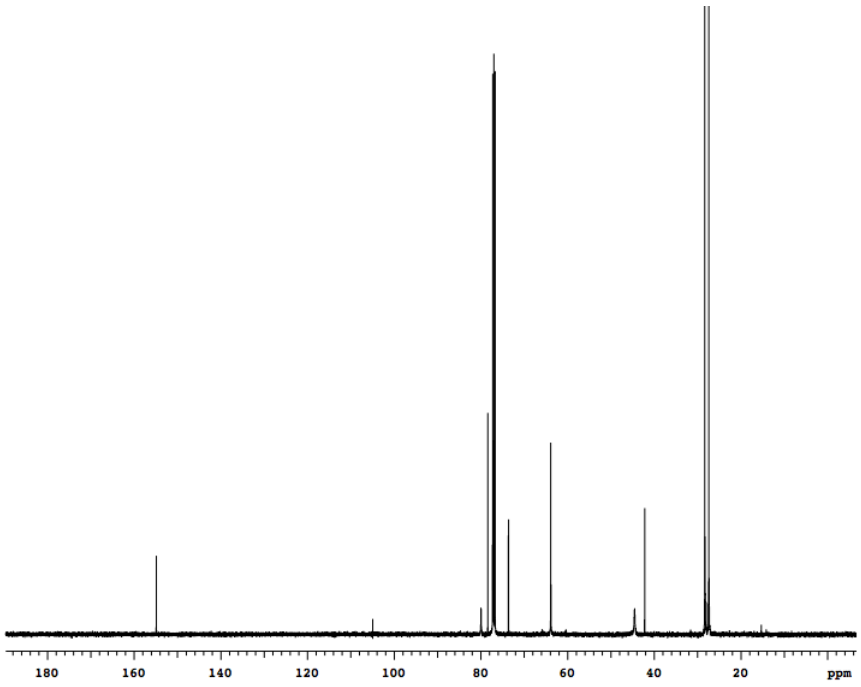
^{1}H NMR (500 MHz, $CDCl_3$) δ 5.03 (br, 1H), 4.49 (br, 1H), 3.49 (br, 2H), 2.24 (s, 1H), 1.45 (s, 9H), 1.20 (s, 9H

^{13}C NMR (125 MHz, $CDCl_3$) δ 154.93, 82.11, 79.87, 73.50, 70.60, 63.84, 43.38, 28.31, 27.39 (107.25 is noise from NMR.)

MS (ESI, m/z) calcd for $C_{13}H_{24}NO_3$ $(M+H)^+$ 242.18, found 242.18



^{1}H NMR of **8h**

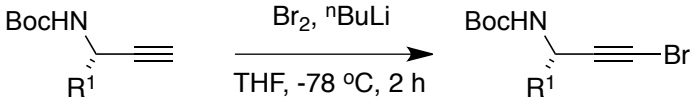


^{13}C NMR of **8h**

General Procedure for Compounds 9a - c

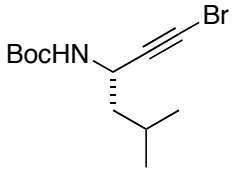
To a solution of the amino acid alkyne (1.0 equiv) in anhydrous THF (0.37 M) was added ^{n}BuLi (2.0 equiv, 2.5 M in hexane) at -78 ^{o}C under N_2 (g). After stirring for 1h, bromine (1.0 equiv) was added dropwise to the solution of the lithium acetylide. The reaction mixture was stirred at -78 ^{o}C under N_2 (g) for 2 h. The mixture was then quenched by adding saturated $Na_2S_2O_3$ (aq.) and allowed to heat up to 25 ^{o}C. The reaction mixture was extracted with diethyl ether. The combined ether fractions were washed with brine, dried over $MgSO_4$ and then concentrated under vacuum. The compounds **9a - c** were purified by Flash chromatography.

Scheme S5. Synthesis of compounds **9a - c**.



8a, c, h **9a-c**

Preparation of Compound 9a

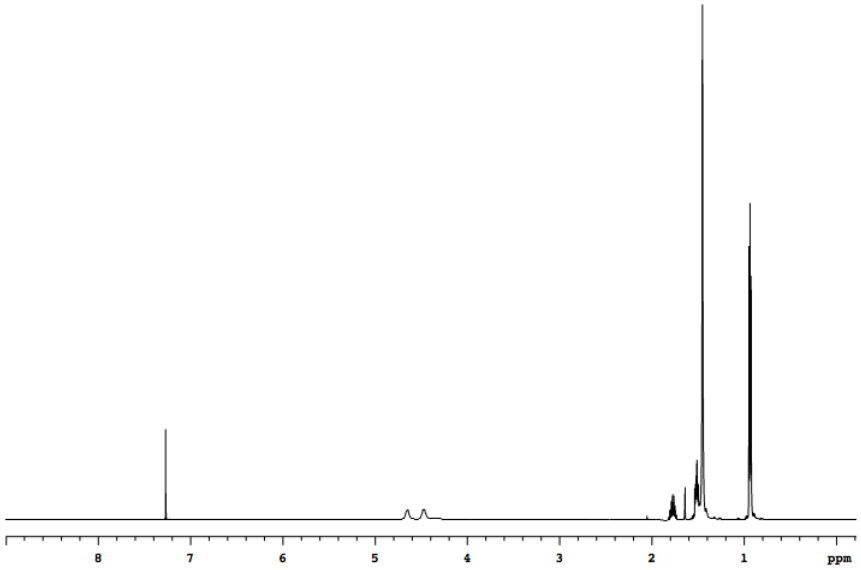


9a

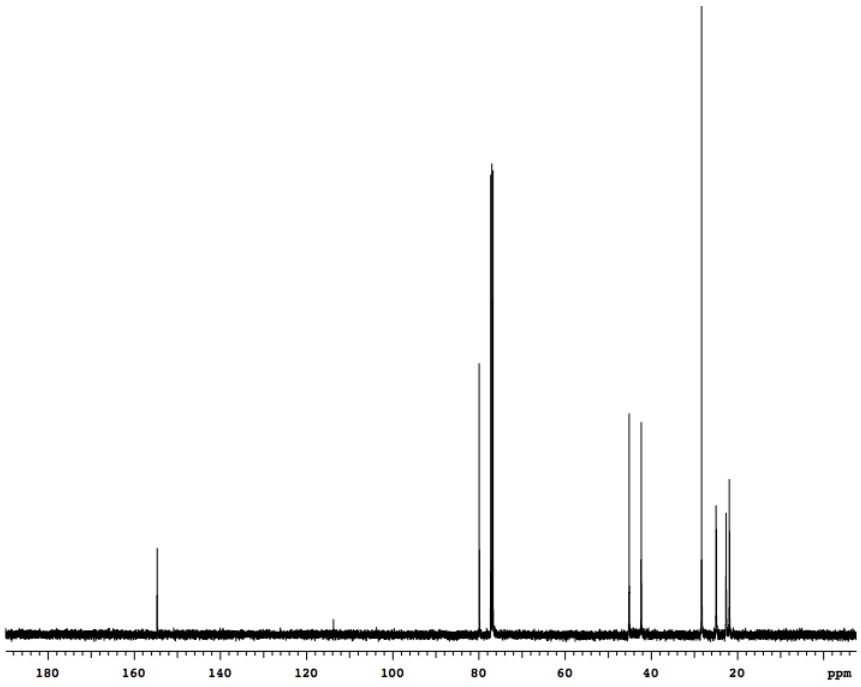
Compound **9a** was prepared from **8a** (1.50 g, 7.1 mmol). Flash chromatography (3:97 EtOAc/Hexanes) afforded 2.01 g (98 %) **9a** as a white solid.

^{1}H NMR (500 MHz, $CDCl_3$) δ 4.65 (br, 1H), 4.49-4.46 (m, 1H), 1.82-1.73 (m, 1H), 1.53-1.50 (m, 2H), 1.45 (s, 9H), 0.94 (d, 3H, J = 5.5 Hz), 0.93 (d, 3H, J = 5.5 Hz)

^{13}C NMR (125 MHz, $CDCl_3$) δ 154.72, 113.78, 79.95, 45.16, 42.33, 28.36, 24.99, 22.67, 21.91

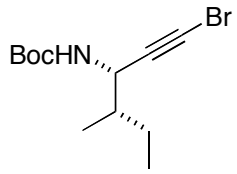


^{1}H NMR of **9a**



^{13}C NMR of **9a**

Preparation of Compound 9b

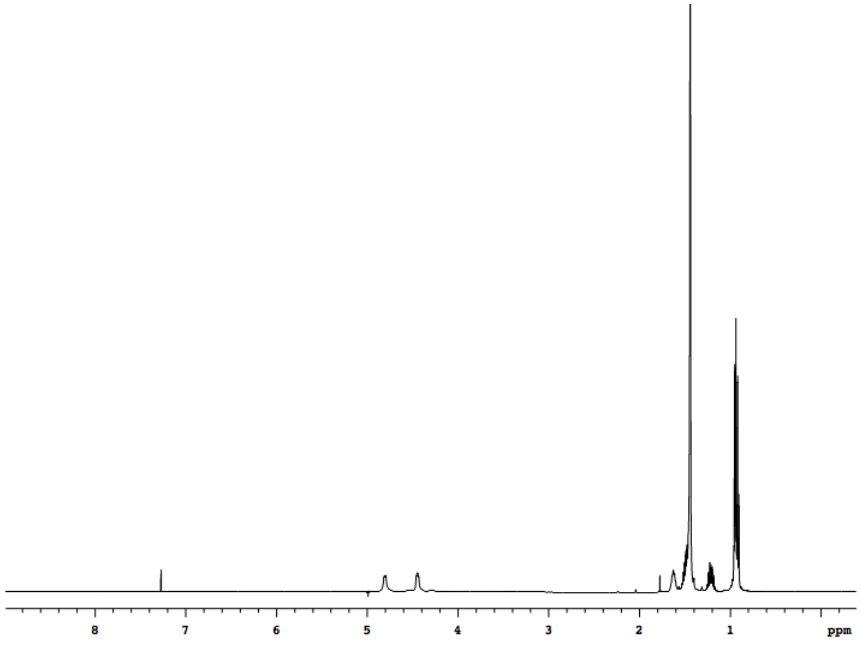


9b

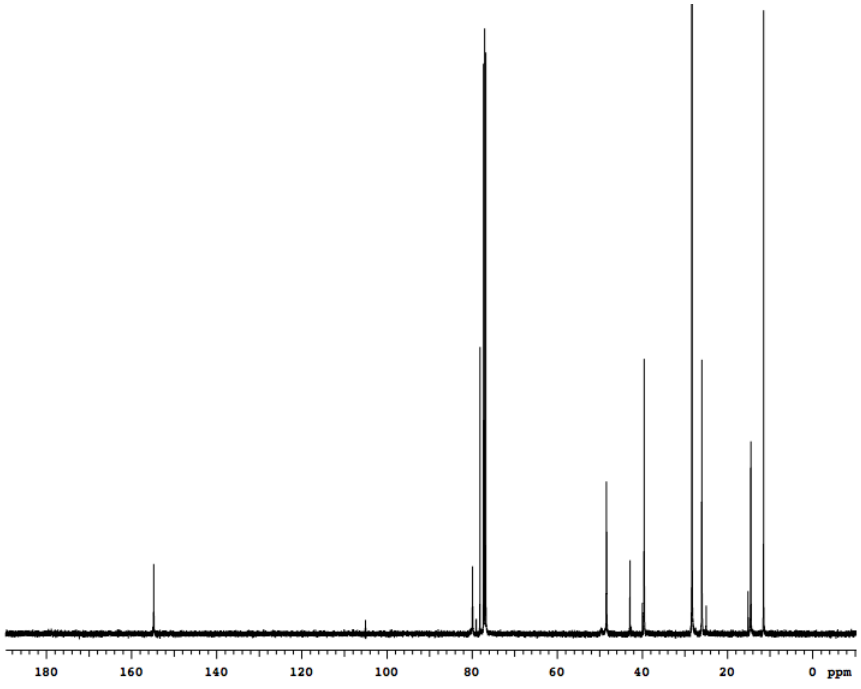
Compound **9b** was prepared from **8c** (0.68 g, 3.2 mmol). Flash chromatography (1:19 EtOAc/Hexanes) afforded 0.88 g (94 %) **9b** as a yellowish solid.

^{1}H NMR (500 MHz, $CDCl_3$) δ 4.80 (d, 1H, J = 3.0 Hz), 4.45 (dd, 1H, J = 5.0, 8.5 Hz), 1.64-1.60 (m, 1H), 1.53-1.46 (m, 1H), 1.44 (s, 9H), 1.26-1.17 (m, 1H), 0.94 (d, 3H, J = 6.5 Hz), 0.92 (t, 3H, J = 8.0 Hz)

^{13}C NMR (125 MHz, $CDCl_3$) δ 154.75, 79.85, 78.11, 48.39, 42.86, 39.54, 28.31, 25.99, 14.48, 11.49 (107.25 is noise from NMR.)

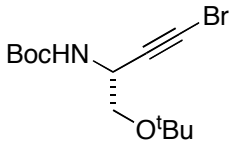


^{1}H NMR of **9b**



^{13}C NMR of **9b**

Preparation of Compound 9c

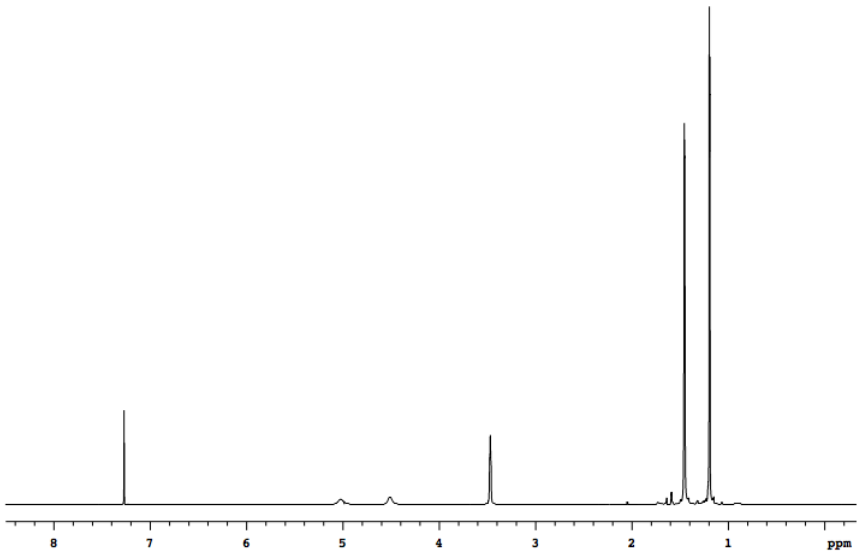


9c

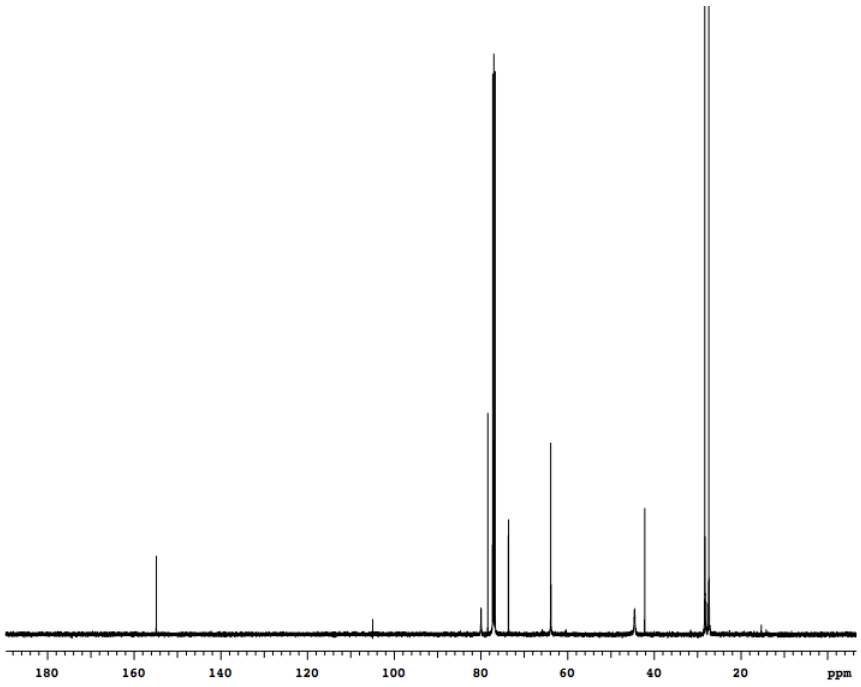
Compound **9c** was prepared from **8h** (0.40 g, 1.7 mmol). Flash chromatography (1:9 EtOAc/Hexanes) afforded 0.45 g (84 %) **9c** as a yellowish solid.

^{1}H NMR (500 MHz, $CDCl_3$) δ 5.02 (br, 1H), 4.51 (br, 1H), 3.47 (br, 2H), 1.45 (s, 9H), 1.20 (s, 9H)

^{13}C NMR (125 MHz, $CDCl_3$) δ 154.86, 79.95, 78.38, 73.61, 63.85, 44.50, 42.18, 28.31, 27.40 (107.25 is noise from NMR.)



^{1}H NMR of **9c**



^{13}C NMR of **9c**

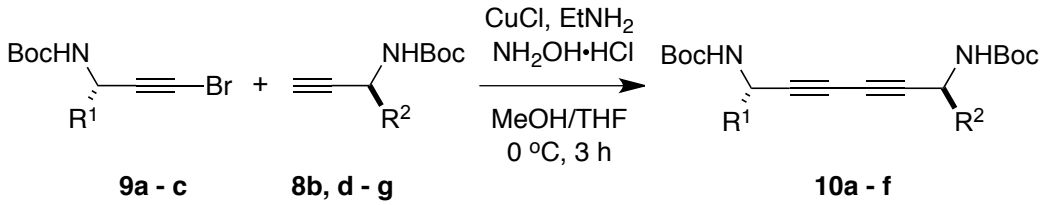
General Procedure for Compounds 10a - f

To a solution of **8b, d - g** (1.0 equiv) in MeOH (0.72 M) were added CuCl (0.1 equiv), $NH_2OH{\cdot}H_2O$ (0.95 equiv) and ethyl amine (25.0 equiv) at 0 ^oC under N_2 (g). A THF solution of **9a - c** (1.2 equiv) was added dropwise to the reaction mixture at 0 ^oC over 1 h. After adding the THF solution, the mixture was stirred at 0 ^oC for 2 h. The mixture was then quenched by added H_2O and extracted with diethyl ether. The combined ether fractions were dried over $MgSO_4$ and then concentrated under vacuum. The compounds **10a - f** were purified by Flash chromatography.

For compound **10d'**, to a solution of **10d** (1.0 equiv) in THF (0.082 M) was added TBAF (1.2 equiv) at 0 ^oC. After stirring at 0 ^oC for 15 min, the reaction mixture was warmed to 25 ^oC and then stirred for 45 min. The reaction mixture was diluted with H_2O and

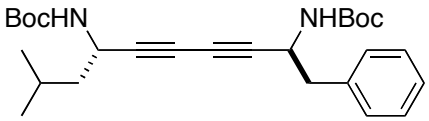
extracted with EtOAc. The combined organic phases were dried over $MgSO_4$. After completely removing the solvent, the **10d'** was purified by flash chromatography.

Scheme S6. Synthesis of compounds **10a - f**.



9a - c **8b, d - g** **10a - f**

Preparation of Compound 10a



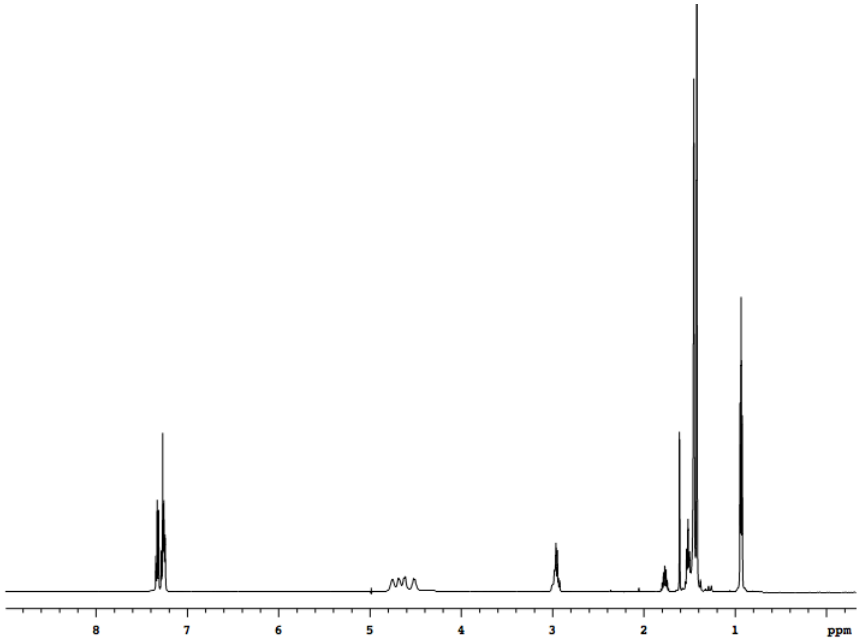
10a

Compound **10a** was prepared from **9a** (0.26 g, 0.88 mmol) and **8b** (0.18 g, 0.73 mmol). Flash chromatography (1:19 to 1:9 EtOAc/Hexanes) afforded 0.31g (94 %) **10a** as a white solid.

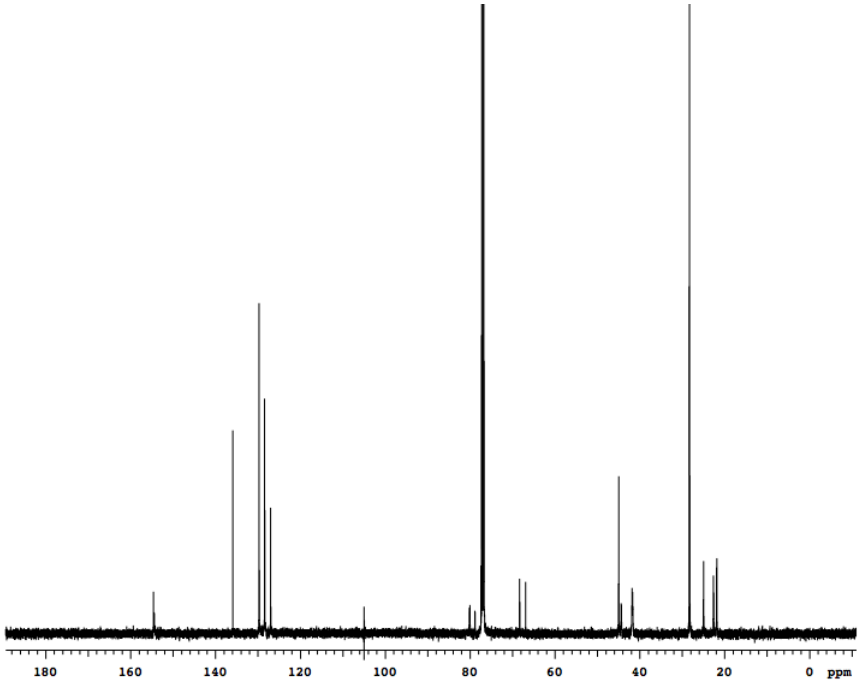
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.33 (t, 2H, J = 7.5 Hz), 7.29-7.24 (m, 3H), 4.76 (br, 1H), 4.69 (br, 1H), 4.62 (br, 1H), 4.52 (br, 1H), 3.00-2.98 (m, 1H), 2.94 (dd, 1H, J = 7.0, 13.5 Hz), 1.81-1.73 (m, 1H), 1.52 (dd, 2H, J = 7.5, 7.5 Hz), 1.45 (s, 9H), 1.42 (s, 9H), 0.94 (d, 3H, J = 6.5 Hz), 0.93 (d, 3H, J = 5.5 Hz)

^{13}C NMR (125 MHz, $CDCl_3$) δ 154.61, 154.44, 135.93, 129.70, 128.41, 126.99, 80.15, 80.04, 78.84, 77.44, 68.31, 66.91, 44.96, 44.37, 41.84, 28.32, 28.27, 24.97, 22.63, 21.85 (107.25 is noise from NMR.)

MS (MALDI, m/z) calcd for $C_{27}H_{38}N_2NaO_4$ $(M+Na)^+$ 477.27, found 477.34

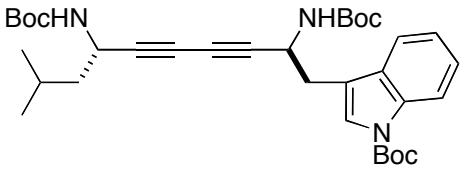


^{1}H NMR of **10a**



^{13}C NMR of **10a**

Preparation of Compound 10b



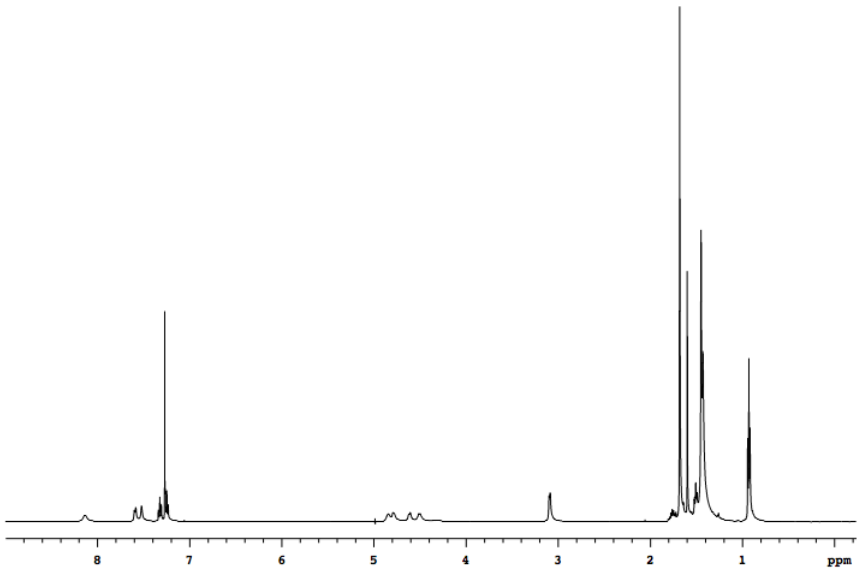
10b

Compound **10a** was prepared from **9a** (0.27 g, 0.94 mmol) and **8f** (0.30 g, 0.78 mmol). Flash chromatography (1:9 EtOAc/Hexanes) afforded 0.39 g (84 %) **10a** as a white solid.

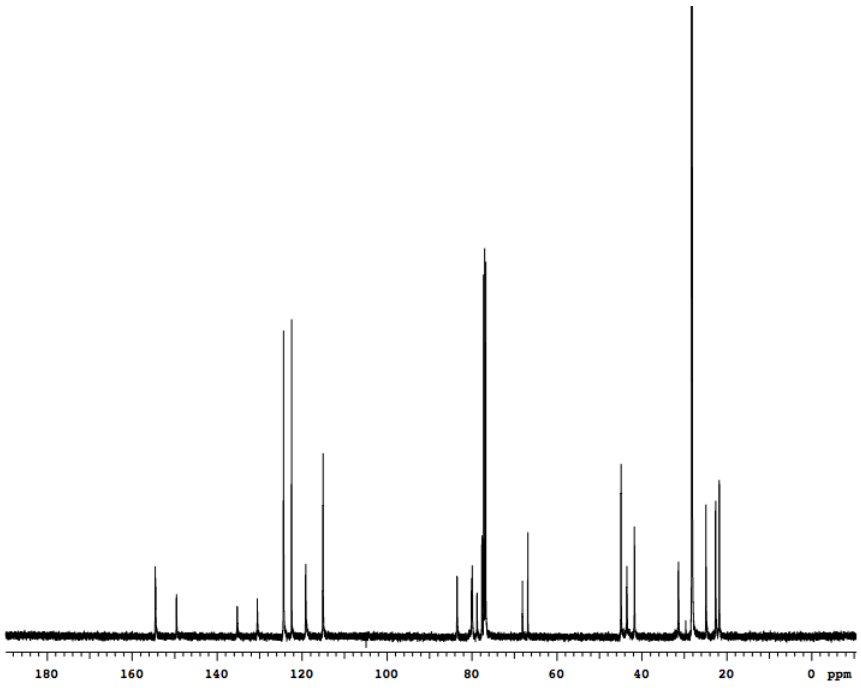
^{1}H NMR (500 MHz, $CDCl_3$) δ 8.13 (br, 1H), 7.59 (d, 1H. J = 8.0 Hz), 7.52 (br, 1H), 7.32 (t, 1H, J = 7.0 Hz), 7.25 (t, 1H, J = 8.0 Hz), 4.85 (br, 1H), 4.79 (br, 1H), 4,61 (br, 1H), 4.50 (br, 1H), 3.09 (d, 2H, 6.0 Hz), 1.79-1.72 (m, 1H), 1.68 (s, 9H), 1.51 (dd, 2H, J = 7.5, 7.5 Hz), 1.45 (s, 9H), 1.43 (s, 9H), 0.94 (d, 3H, 7.0 Hz), 0.92 (d, 3H, 6.5 Hz)

^{13}C NMR (125 MHz, $CDCl_3$) δ 154.56, 154.47, 149.55, 135.25, 130.51, 124.34, 124.34, 122.46, 119.12, 115.09, 115.03, 83.48, 80.06, 79.90, 78.78, 77.63, 68.10, 66.82, 44.89, 43.51, 41.71, 31.36, 28.26, 28.22, 28.13, 24.89, 22.61, 21.74

MS (MALDI, m/z) calcd for $C_{34}H_{47}N_3NaO_6$ $(M+Na)^+$ 616.34, found 616.40

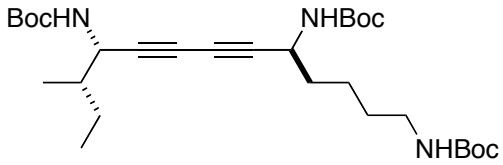


^{1}H NMR of **10b**



^{13}C NMR of **10b**

Preparation of Compound 10c



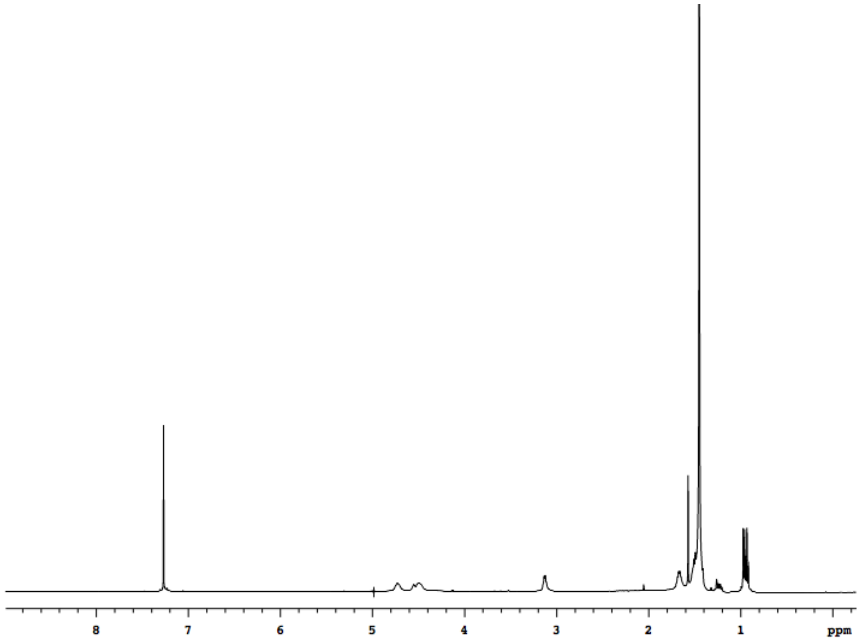
10c

Compound **10c** was prepared from **9b** (0.25 g, 0.85 mmol) and **8d** (0.23 g, 0.70 mmol). Flash chromatography (1:19 EtOAc/Hexanes) afforded 0.27 g (71 %) **10c** as a colorless oil.

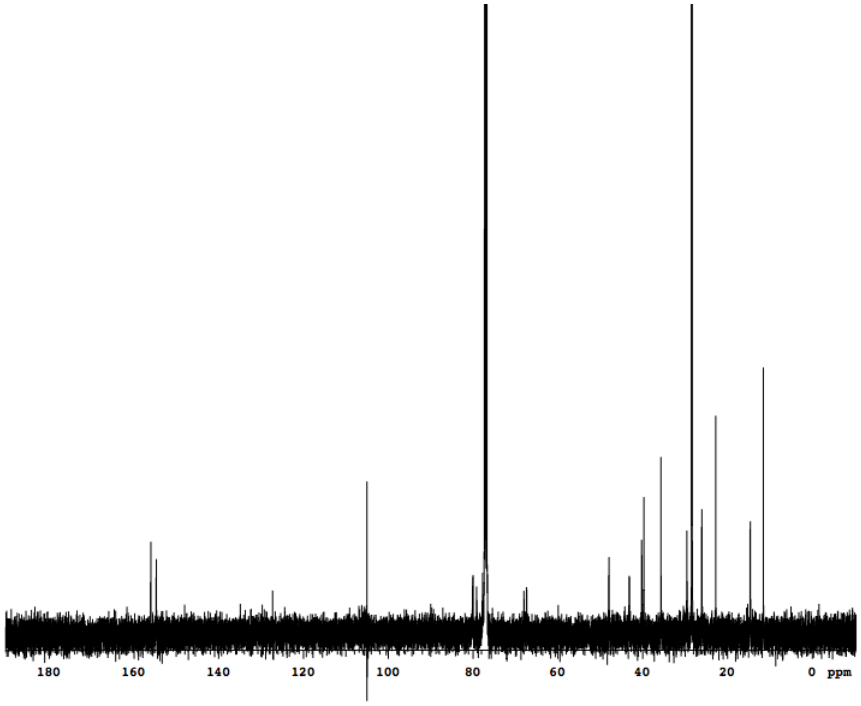
^{1}H NMR (500 MHz, $CDCl_3$) δ 4.73 (br, 2H), 4.55 (br, 1H), 4.49 (br, 2H), 3.14-3.12 (m, 2H), 1.69-1.61 (m, 3H), 1.54-1.45 (m, 5H), 1.28-1.20 (m, 1H), 0.97 (d, 3H, J = 6.5 Hz), 0.92 (t, 3H, J = 7.5 Hz)

^{13}C NMR (125 MHz, $CDCl_3$) δ 156.02, 154.71, 154.71, 80.12, 80.01, 80.01, 79.14, 77.71, 67.97, 67.36, 47.95, 43.14, 40.21, 39.70, 35.66, 29.53, 28.42, 28.33, 28.33, 26.04, 22.74, 14.60, 11.50 (107.25 is noise from NMR.)

MS (MALDI, m/z) calcd for $C_{29}H_{49}N_3NaO_6$ $(M+Na)^+$ 558.35, found 558.17

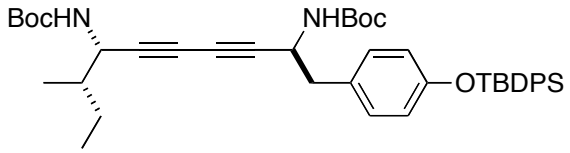


^{1}H NMR of **10c**



^{13}C NMR of **10c**

Preparation of Compound 10d



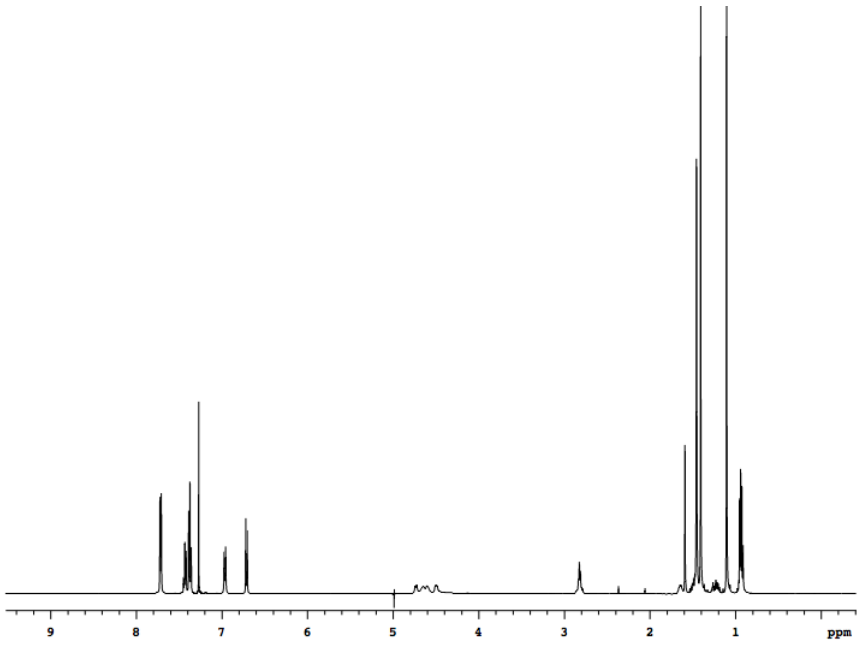
10d

Compound **10d** was prepared from **9b** (0.12 g, 0.41 mmol) and **8e** (0.17 g, 0.34 mmol). Flash chromatography (1:19 EtOAc/Hexanes) afforded 0.18 g (74 %) **10d** as a yellowish solid.

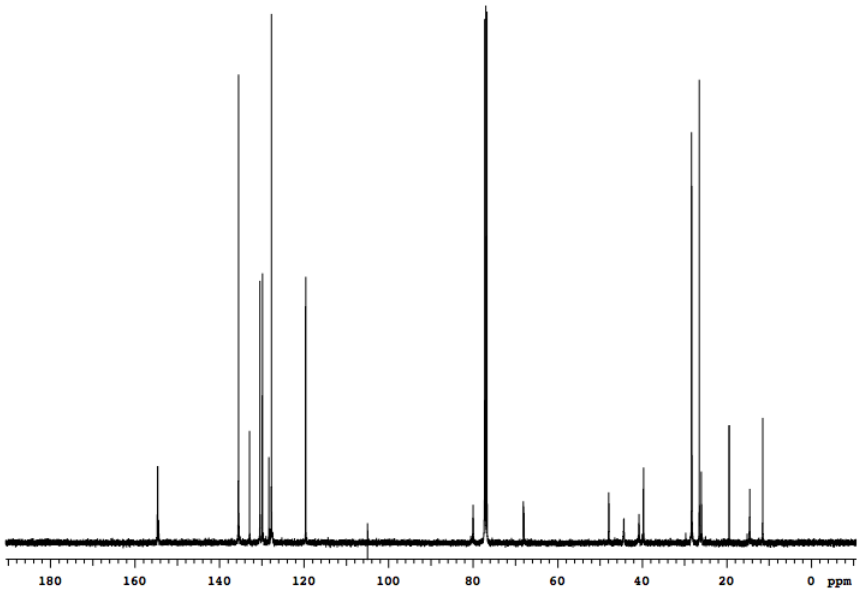
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.72 (d, 4H, J = 7.0 Hz), 7.43 (dd, 2H, J = 7.5, 7.5 Hz), 7.37 (dd, 4H, J = 7.5, 7.5 Hz), 6.97 (d, 2H, J = 8.5 Hz), 6.71 (d, 2H, J = 9.0 Hz), 4.73 (d, 1H, J = 8.5 Hz), 4.65 (br, 1H), 4.61 (br, 1H), 4.50 (br, 1H), 2.88-2.83 (m, 1H), 2.81 (dd, 1H, J = 6.5, 13.5 Hz), 1.68-1.62 (m, 1H), 1.53-1.46 (m, 1H), 1.46 (s, 9H), 1.41 (s, 9H), 1.27-1.19 (m, 1H), 1.10 (s, 9H), 0.95 (d, 3H, J = 6.0 Hz), 0.93 (t, 3H, J = 7.5 Hz)

^{13}C NMR (125 MHz, $CDCl_3$) δ 154.71, 154.62, 154.41, 135.49, 132.89, 130.42, 129.84, 128.31, 127.71, 119.61, 79.99, 79.99, 77.34, 77.34, 68.09, 67.97, 47.90, 44.34, 40.76, 39.68, 28.31, 28.26, 26.49, 26.00, 19.3, 14.56, 11.49 (107.25 is noise from NMR.)

MS (MALDI, m/z) calcd for $C_{43}H_{56}N_2NaO_5Si$ $(M+Na)^+$ 731.39, found 731.40

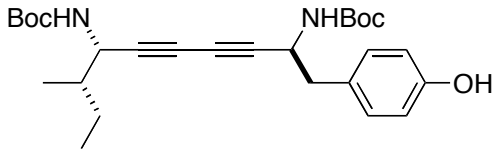


^{1}H NMR of **10d**



^{13}C NMR of **10d**

Preparation of Compound 10d'



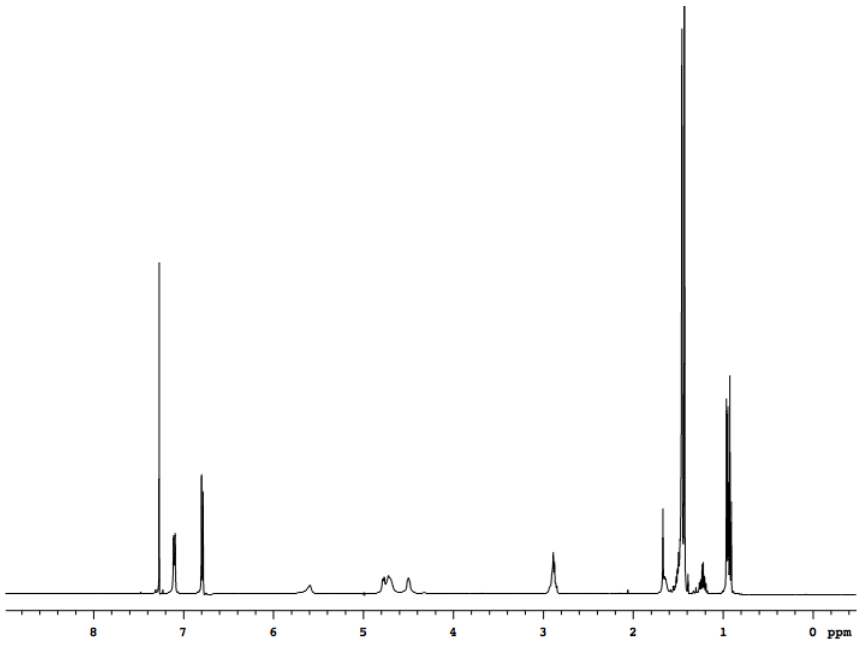
10d'

Compound **10d'** was prepared from **10d** (94.50 mg, 0.13 mmol). Flash chromatography (1:5 EtOAc/Hexanes) afforded (57.10 mg, 91 %) **10d'** as a yellowish solid.

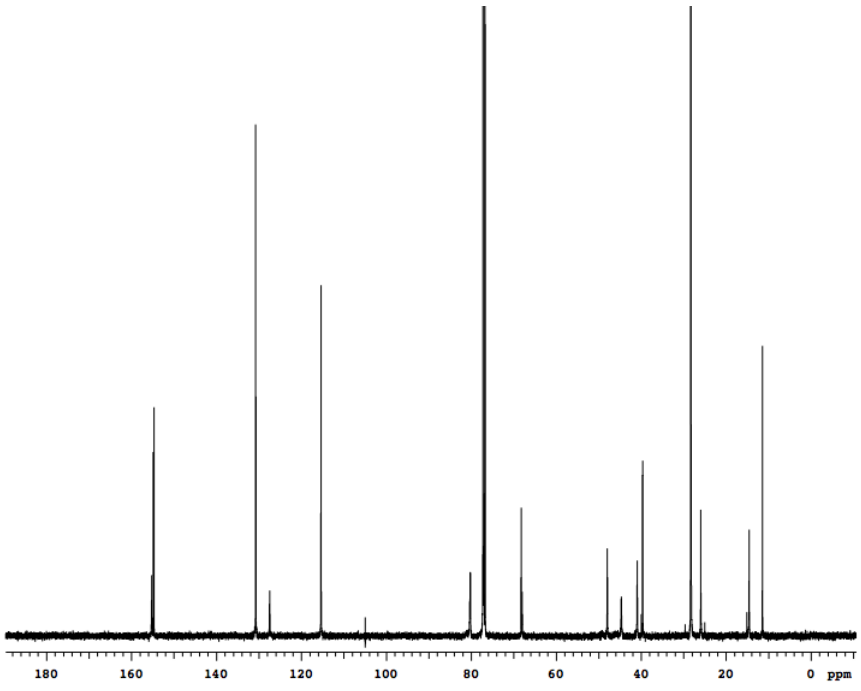
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.11 (d, 2H, J = 8.0 Hz), 6.79 (d, 2H, J = 8.5 Hz), 5.59 (br, 1H), 4.78 (d, 1H, J = 8.0 Hz), 4.72 (br, 2H), 4.50 (br, 1H), 2.94-2.89 (m, 1H), 2.87 (dd, 1H, J = 6.5, 13.0 Hz), 1.67-1.64 (m, 1H), 1.54-1.46 (m, 1H), 1.46 (s, 9H), 1.43 (s, 9H), 1.27-1.18 (m, 1H), 0.96 (d, 3H, J = 7.0 Hz), 0.92 (d, 3H, J = 7.5 Hz)

^{13}C NMR (125 MHz, $CDCl_3$) δ 155.23, 154.89, 154.69, 130.75, 127.45, 115.37, 80.34, 80.21, 77.13, 77.13, 68.24, 68.00, 47.96, 44.68, 40.91, 39.65, 28.31, 28.27, 25.97, 14.57, 11.45 (107.25 is noise from NMR.)

MS (MALDI, m/z) calcd for $C_{27}H_{38}N_2NaO_5$ $(M+Na)^+$ 493.27, found 493.33

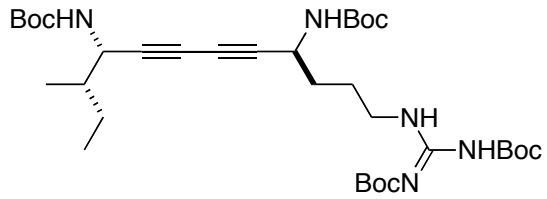


^{1}H NMR of **10d'**



^{13}C NMR of **10d'**

Preparation of Compound 10e



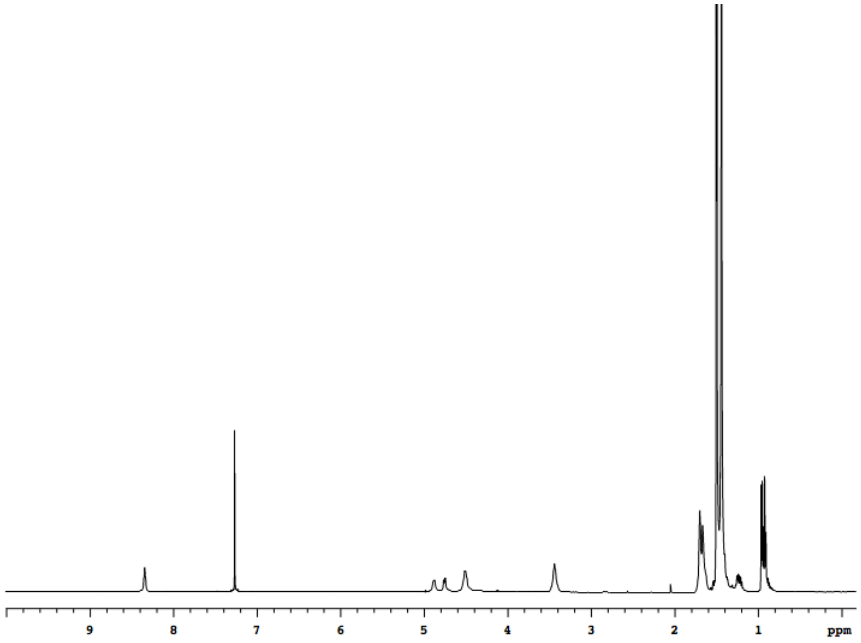
10e

Compound **10e** was prepared from **9b** (0.11 g, 0.4 mmol) and **8g** (0.15 g, 0.33 mmol). Flash chromatography (1:19 to 1:5 % EtOAc/Hexanes) afforded 0.21 g (95 %) **10e** as a yellowish solid.

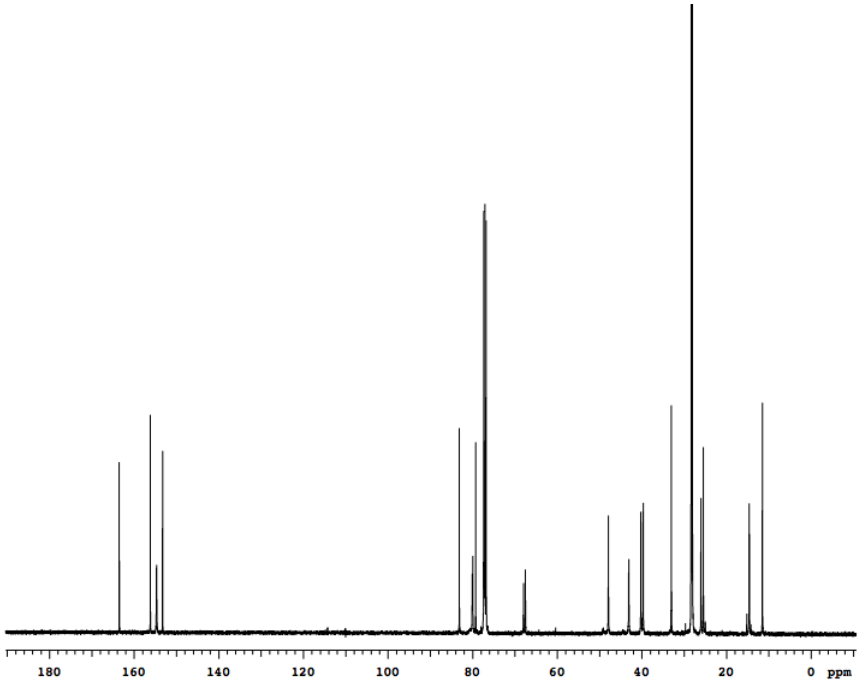
^{1}H NMR (500 MHz, $CDCl_3$) δ 8.35 (t, 1H, J = 4.5 Hz), 4.88 (d, 1H, J = 7.5 Hz), 4.76 (d, 1H, J = 13.5 Hz), 4.51 (br, 2H), 3.44 (br, 2H), 1.71-1.67 (m, 7H), 1.54-1.51 (m, 1H), 1.51 (s, 9H), 1.50 (s, 9H), 1.44 (s, 18H), 1.30-1.15 (m, 1H), 0.96 (d, 3H, J = 6.5 Hz), 0.93 (d, 1H, J = 7.5 Hz)

^{13}C NMR (125 MHz, $CDCl_3$) δ 163.42, 156.08, 154.66, 154.58, 153.18, 83.05, 83.05, 79.99, 79.88, 79.18, 79.18, 67.86, 67.43, 47.82, 42.99, 40.14, 39.59, 32.94, 28.24, 28.18, 27.97, 27.97, 25.94, 25.38, 14.51, 11.43

MS (MALDI, m/z) calcd for $C_{34}H_{57}N_5NaO_8$ $(M+Na)^+$ 686.41, found 686.43

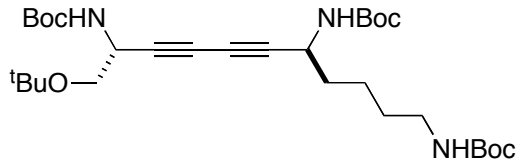


^{1}H NMR of **10e**



^{13}C NMR of **10e**

Preparation of Compound 10f



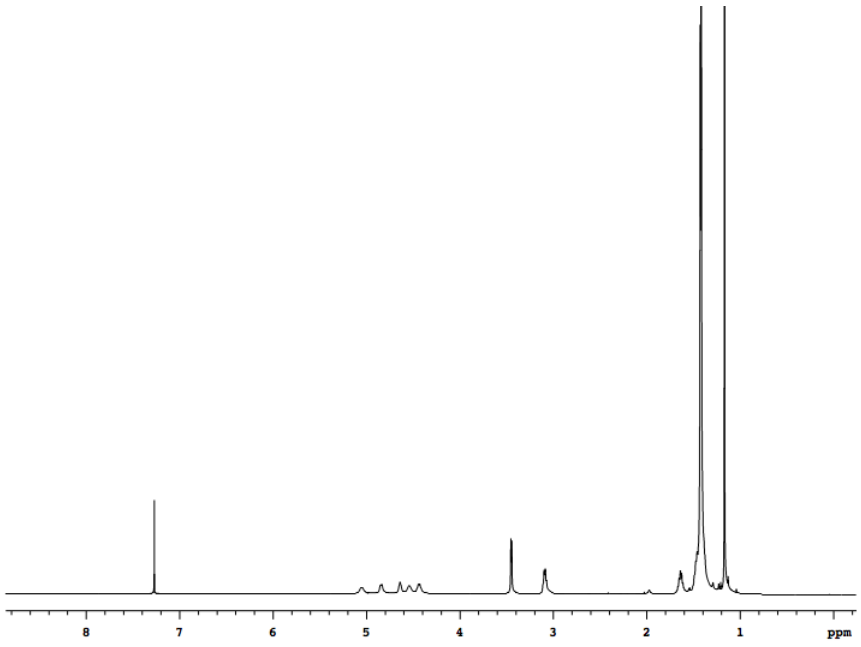
10f

Compound **10f** was prepared from **9c** (0.16 g, 0.51 mmol) and **8d** (0.14 g, 0.43 mmol). Flash chromatography (1:19 EtOAc/Hexanes) afforded 0.18 g (73 %) **10f** as oil.

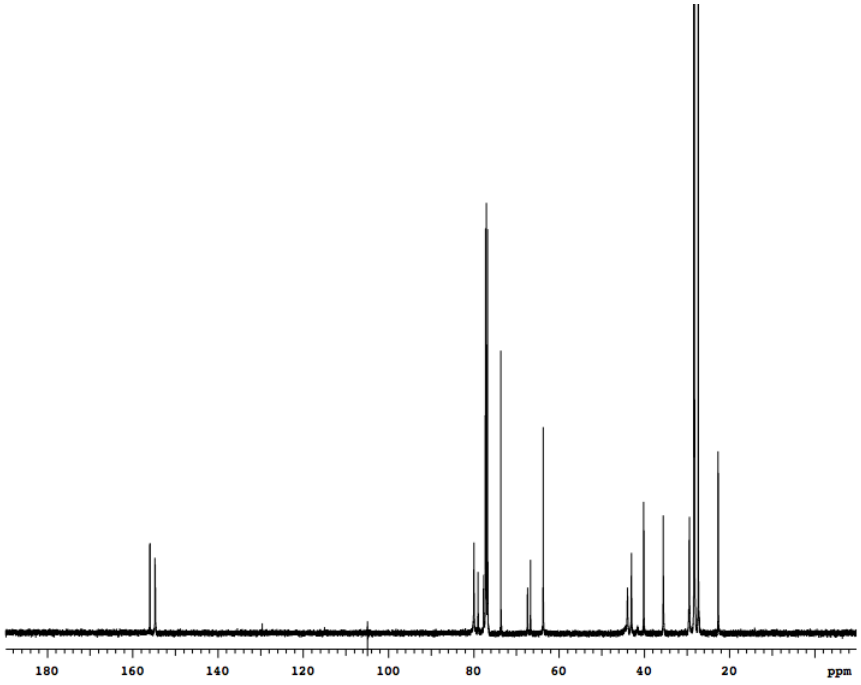
^{1}H NMR (500 MHz, $CDCl_3$) δ 5.05 (br, 1H), 4.84 (d, 1H, J = 8.0 Hz), 4.64 (br, 1H), 4.54 (br, 1H), 4.46-4.40 (m, 1H), 3.45 (d, 2H, J = 4.5 Hz), 3.11-3.07 (m, 1H), 1.67-1.59 (m, 2H), 1.49-1.42 (br, 31H), 1.17 (s, 9H)

^{13}C NMR (125 MHz, $CDCl_3$) δ 155.96, 154.74, 154.69, 79.95, 78.99, 78.99, 77.66, 77.11, 73.63, 67.38, 66.68, 63.71, 43.93, 43.00, 40.13, 35.54, 29.42, 28.35, 28.25, 28.25, 27.33, 22.66 (107.25 is noise from NMR.)

MS (MALDI, m/z) calcd for $C_{30}H_{51}N_3NaO_7$ $(M+Na)^+$ 588.36, found 588.41



^{1}H NMR of **10f**

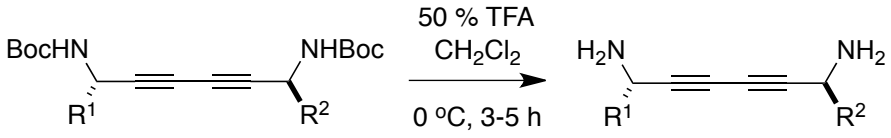


^{13}C NMR of **10f**

General Procedure for Compounds 4a - f

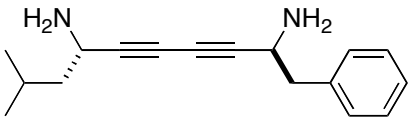
The compounds **10a - f** were treated with 50% TFA/CH_2Cl_2 at 0 ^{o}C for 3-5 h. Evaporating solvent under nitrogen afforded the desired product **4a - f**.

Scheme S7. Synthesis of compounds **4a - f**.



10a - f **4a-f**

Preparation of Compound 4a



4a

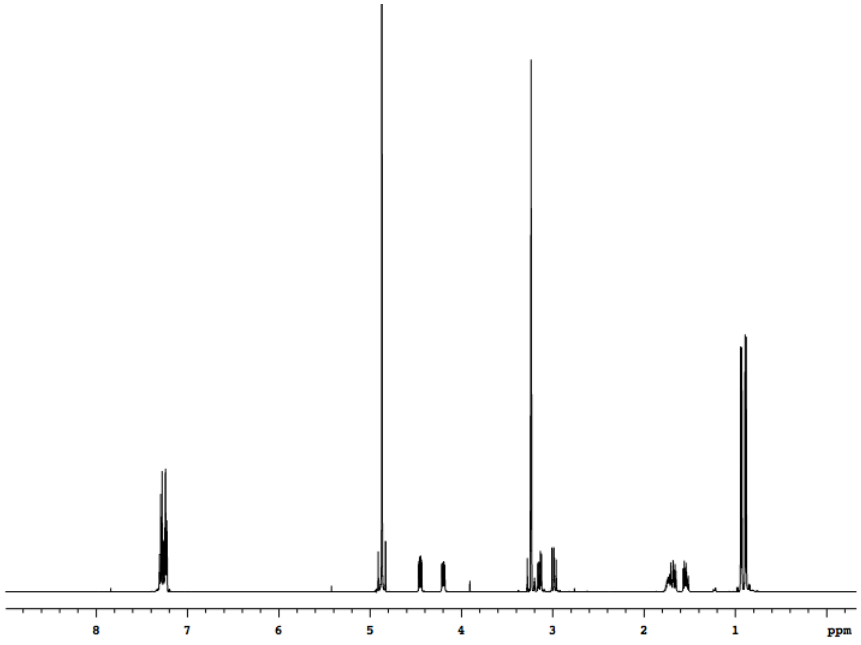
Compound **4a** was prepared from **10a** (10.0 mg, 0.022 mmol). **4a** was obtained as a yellowish amorphous solid (90 %, 2TFA salt).

^{1}H NMR (500 MHz, CD_3OD) δ 7.31-7.23 (m, 5H), 4.45 (dd, 1H, J = 5.5, 9.5 Hz), 4.20 (dd, 1H, J = 5.5, 10.5 Hz), 3.15 (dd, 1H, J = 5.5, 13.5 Hz), 2.99 (dd, 1H, J = 9.5, 13.5 Hz), 1.76-1.71 (m, 1H), 1.71-1.65 (m, 1H), 1.57-1.52 (m, 1H), 0.94 (d, 3H, J = 6.5 Hz), 0.89 (d, 3H, J = 6.5 Hz)

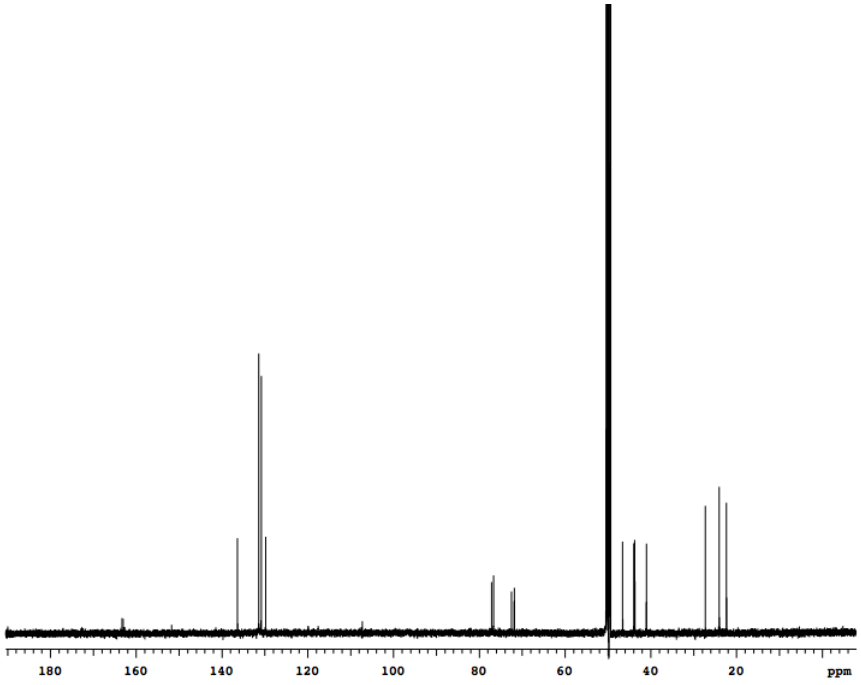
^{13}C NMR (125 MHz, CD_3OD) δ 136.36, 131.40, 130.81, 129.80, 77.11, 76.67, 72.48, 71.82, 46.55, 43.95, 43.71, 41.01, 27.26, 24.01, 22.32 (107.25 is noise from NMR.)

MS (ESI, m/z) calcd for $C_{17}H_{23}N_2$ $(M+H)^+$ 255.19, found 255.21

optical rotation $[\alpha]_D^{19.6}$ = +63.7 (*c* = 0.96 in MeOH)

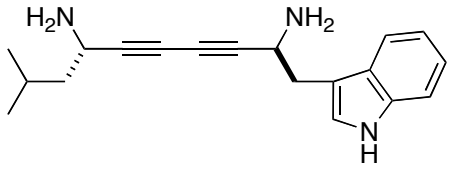


^{1}H NMR of **4a**



^{13}C NMR of **4a**

Preparation of Compound 4b



4b

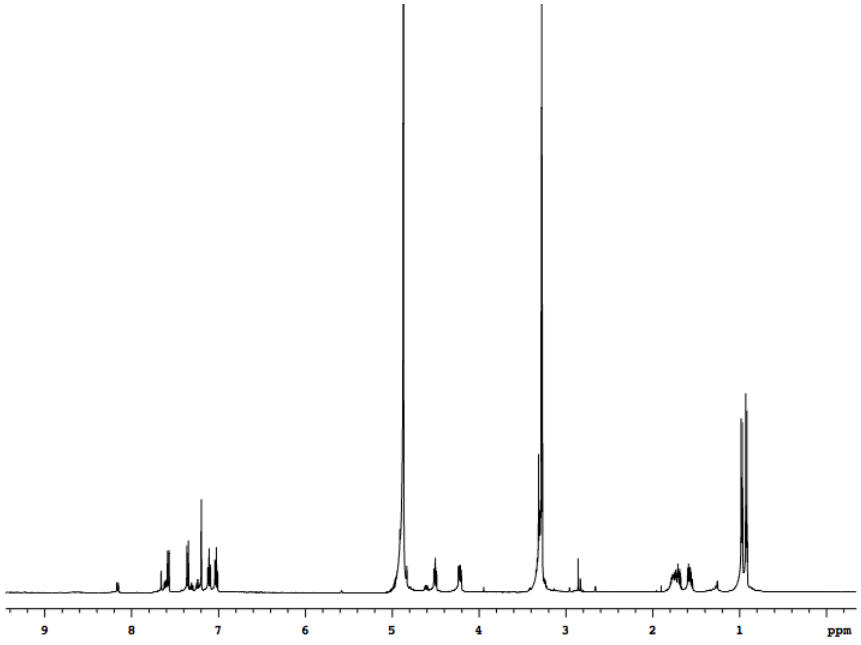
Compound **4b** was prepared from **10b** (9.50 mg, 0.016 mmol). **4b** was obtained as a yellowish amorphous solid (87 %, 2TFA salt).

^{1}H NMR (500 MHz, CD_3OD) δ 7.57 (d, 1H, J = 8.0 Hz), 7.35 (d, 1H. J = 8.5 Hz), 7.20 (s, 1H), 7.10 (dd, 1H, J = 9.5, 9.5 Hz), 7.02 (dd, 1H, J = 8.0 Hz), 4.51 (dd, 1H, 7.5, 8.5 Hz), 4.22 (dd, 1H, J = 5.0, 10.5 Hz), 3.31-3.29 (m, 2H), 1.79-1.74 (m, 1H), 1.74-1,67 (m, 1H), 1.60-1.54 (m, 1H), 0.97 (d, 3H, J = 9.0 Hz), 0.92 (d, 3H, J = 6.5 Hz)

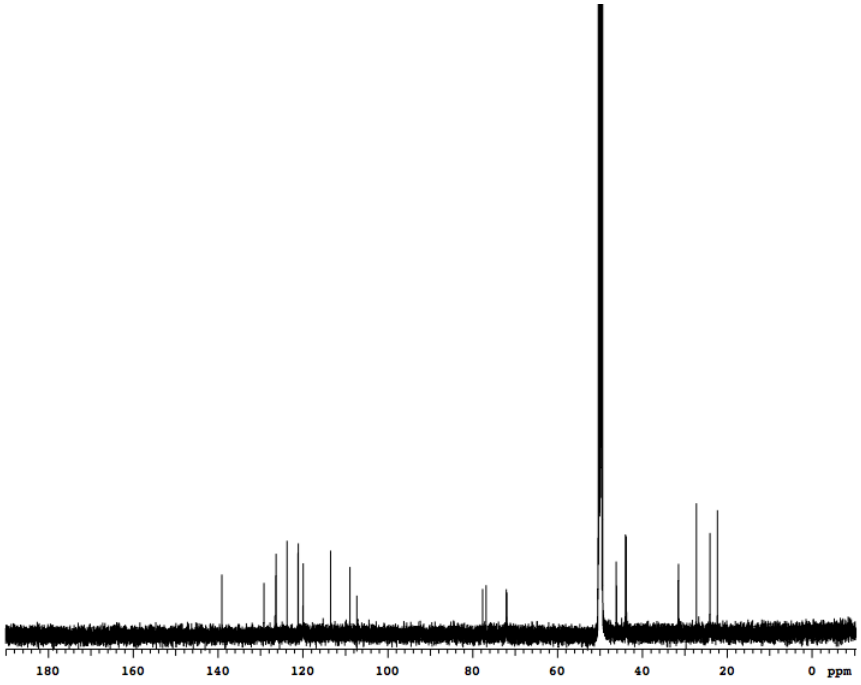
^{13}C NMR (125 MHz, CD_3OD) δ 136.79, 126.89, 124.07, 121.46, 118.83, 117.66, 111.18, 106.63, 75.34, 74.52, 69.79, 69.66, 43.85, 41.69, 41.48, 29.22, 24.95, 21.78, 20.00 (107.25 is noise from NMR.)

MS (ESI, m/z) calcd for $C_{19}H_{24}N_3$ $(M+H)^+$ 294.20, found 294.21

optical rotation $[\alpha]_D^{18.3}$ = +34.8 (*c* = 0.72 in MeOH)

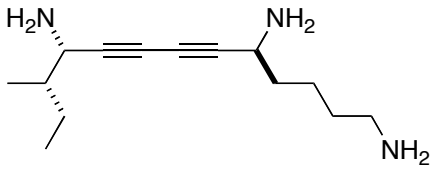


^{1}H NMR of **4b**



^{13}C NMR of **4b**

Preparation of Compound 4c



4c

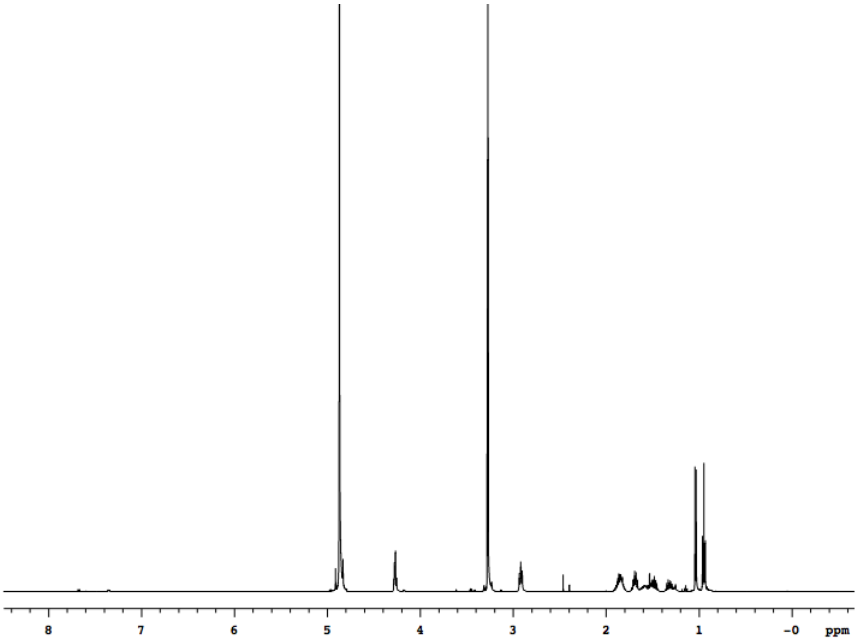
Compound **4c** was prepared from **10c** (4.80 mg, 0.009 mmol). **4c** was obtained as a yellowish amorphous solid (quant., 3TFA salt).

^{1}H NMR (500 MHz, CD_3OD) δ 4.28 (d, 2H, J = 3.5 Hz), 4.27 (d, 2H, J = 10.5 Hz), 2.92 (t, 2H, J = 8.0 Hz), 1.90-1.80 (m, 3H), 1.73-1.66 (m, 2H), 1.53-1.44 (m, 2H), 1.35-1.27 (m, 2H), 1.04 (d, 3H, J = 6.5 Hz), 0.95 (t, 3H, J = 10.5 Hz)

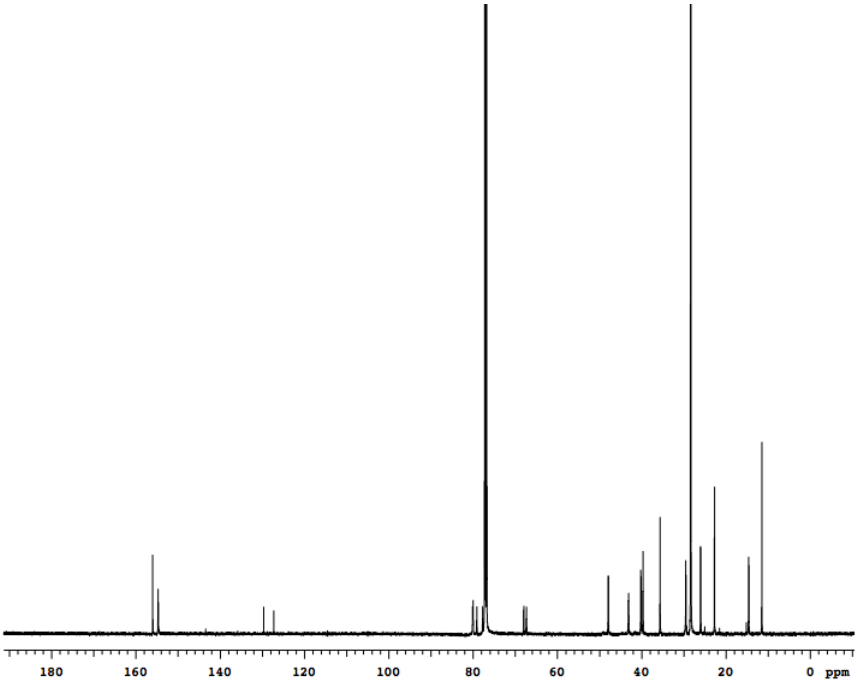
^{13}C NMR (125 MHz, CD_3OD) δ 76.31, 75.45, 72.55, 71.85, 44.89, 40.95, 39.75, 34.42, 28.56, 28.15, 24.26, 24.26, 14.81, 12.22

MS (ESI, m/z) calcd for $C_{14}H_{26}N_3$ $(M+H)^+$ 236.21, found 236.22

optical rotation $[\alpha]_D^{19.8}$ = +32.6 (*c* = 0.21 in MeOH)

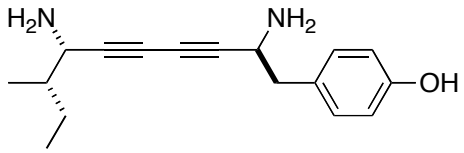


^{1}H NMR of **4c**



^{13}C NMR of **4c**

Preparation of Compound 4d



4d

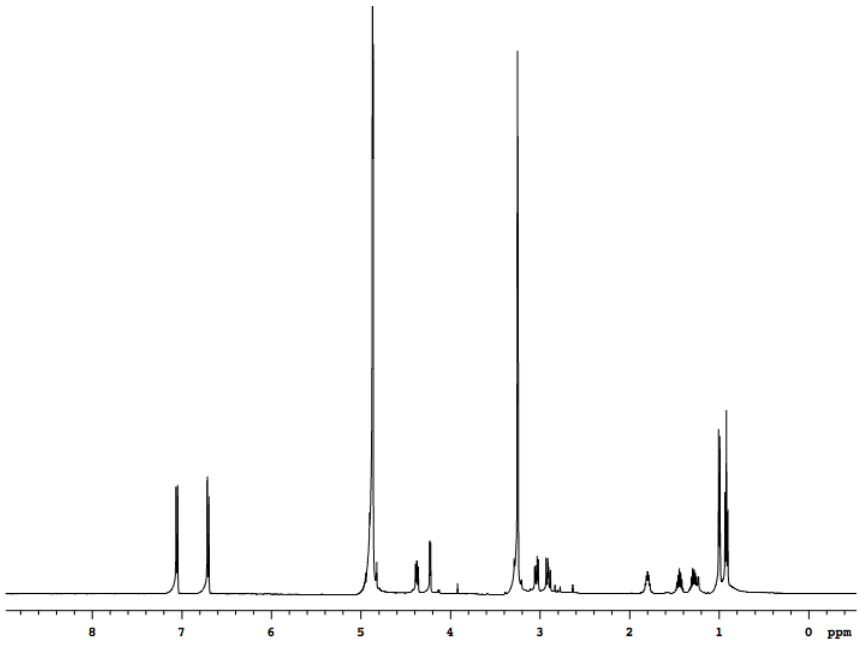
Compound **4d** was prepared from **10d** (10.0 mg, 0.013 mmol). **4d** was obtained as a yellowish amorphous solid (93 %, 2TFA salt).

^{1}H NMR (500 MHz, CD_3OD) δ 7.05 (d, 2H, J = 8.5 Hz), 6.71 (d, 2H, J = 8.5 Hz), 4.38 (dd, 1H, J = 5.5, 9.0 Hz), 4.23 (d, 1H, J = 5.0 Hz), 3.04 (dd, 1H, J = 5.5, 13.5 Hz), 2.91 (dd, 1H, J = 9.5, 13.5 Hz), 1.83-1.78 (m, 1H), 1.83-1.78 (m, 1H), 1.47 (m, 1H), 1.30 (m, 1H), 1.00 (d, 3H, J = 7.0 Hz), 0.92 (t, 3H, J = 7.5 Hz)

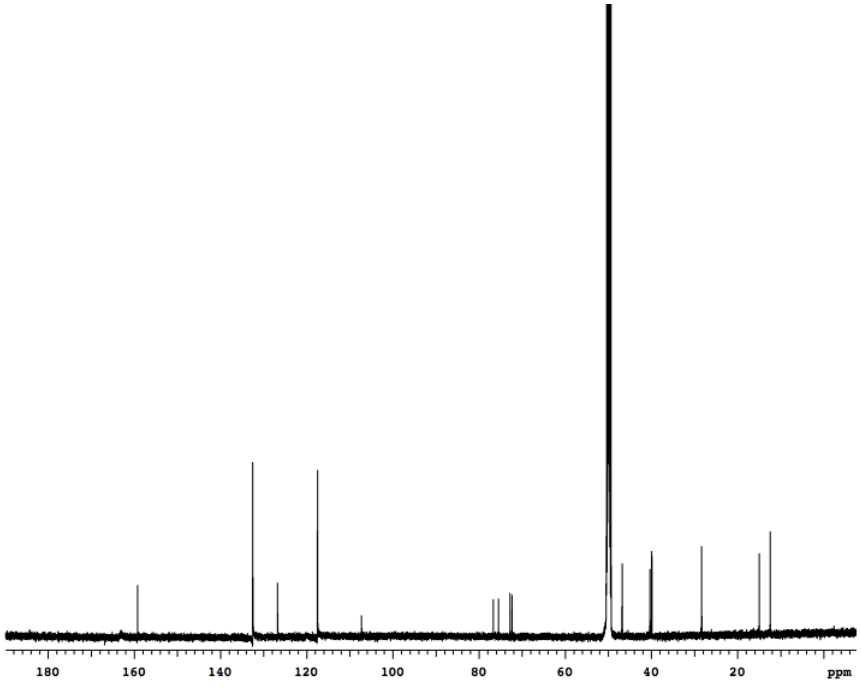
^{13}C NMR (125 MHz, CD_3OD) δ 156.98, 130.21, 124.46, 115.20, 74.40, 73.19, 70.50, 70.08, 44.49, 44.49, 38.03, 37.61, 26.03, 12.66, 10.11 (107.25 is noise from NMR.)

MS (ESI, m/z) calcd for $C_{17}H_{23}N_2O$ $(M+H)^+$ 271.18, found 271.19

optical rotation $[\alpha]_D^{18.6}$ = +40.2 (*c* = 0.60 in MeOH)

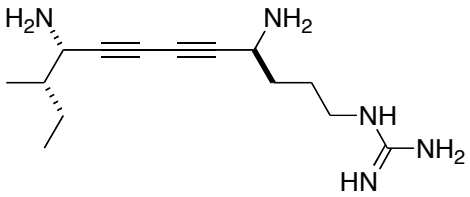


^{1}H NMR of **4d**



^{13}C NMR of **4d**

Preparation of Compound 4e



4e

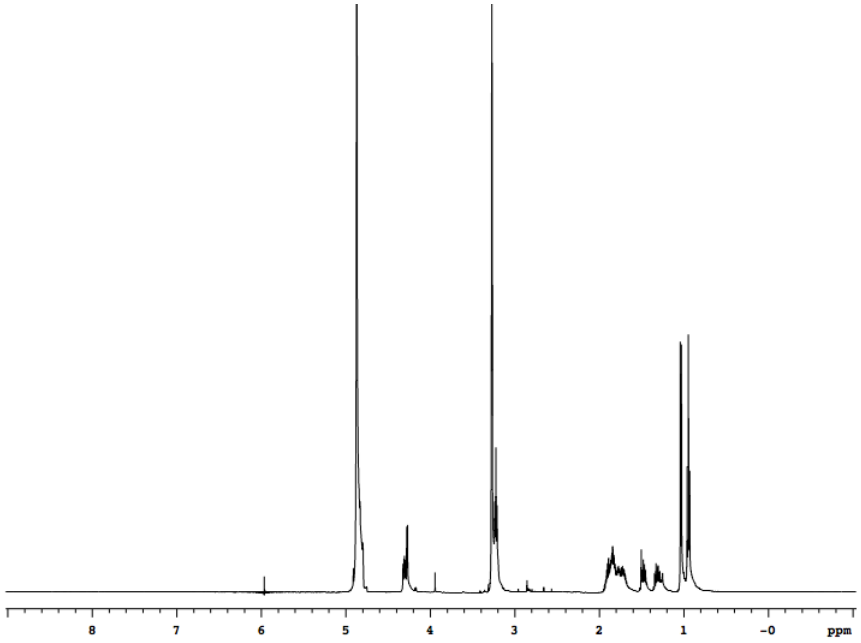
Compound **4e** was prepared from **10e** (9.80 mg, 0.015 mmol). **4e** was obtained as a yellowish amorphous solid (93 %, 3TFA salt).

^{1}H NMR (500 MHz, CD_3OD) δ 4.31 (dd, 1H, J = 5.5, 9.0 Hz), 4.27 (d, 1H, J = 4.5 Hz), 3.22 (t, 2H, J = 7.0 Hz), 1.94-1.87 (m, 1H), 1.86-1.80 (m, 2H), 1.79-1.66 (m, 2H), 1.50-1.44 (m, 1H), 1.36-1.25 (m, 1H), 1.03 (d, 3H, J = 6.5 Hz), 0.95 (t, 3H, J = 10.5)

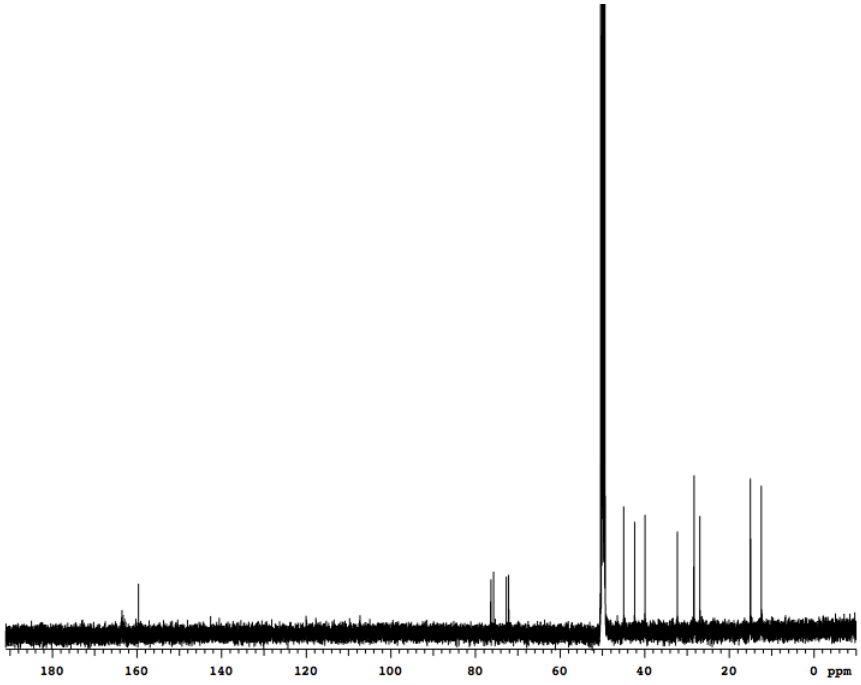
^{13}C NMR (125 MHz, CD_3OD) δ 157.32, 74.04, 73.40, 70.39, 69.82, 42.65, 40.07, 37.62, 37.62, 29.96, 26.04, 24.63, 12.68, 10.11

MS (ESI, m/z) calcd for $C_{14}H_{26}N_5$ $(M+H)^+$ 264.39, found 264.23

optical rotation $[\alpha]_D^{19.4}$ = +8.7 (*c* = 0.84 in MeOH)

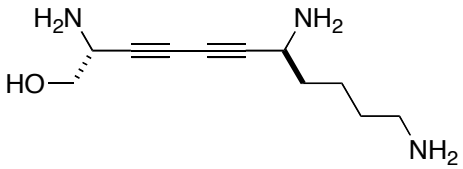


^{1}H NMR of **4e**



^{13}C NMR of **4e**

Preparation of Compound 4f



4f

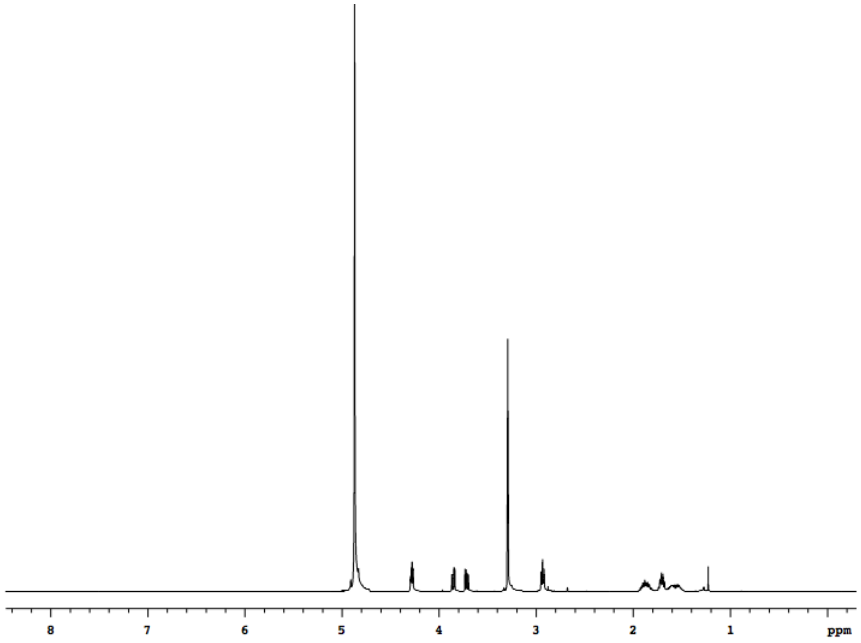
Compound **4f** was prepared from **10f** (10.30 mg, 0.018 mmol). **4f** was obtained as a yellowish amorphous solid (90%, 3TFA salt).

^{1}H NMR (500 MHz, CD_3OD) δ 4.24 (dd, 1H, J = 1.5, 2.0 Hz), 4.22 (d, 1H, J = 5.0 Hz), 3.80 (dd, 1H, J = 4.5, 11.5 Hz), 3.66 (dd, 1H, J = 6.5, 11.5 Hz), 2.88 (t, 2H, J = 7.5 Hz), 1.87-1.75 (m, 2H), 1.68-1.62 (m, 2H), 1.58-1.52 (m, 1H), 1.50-1.45 (m, 1H)

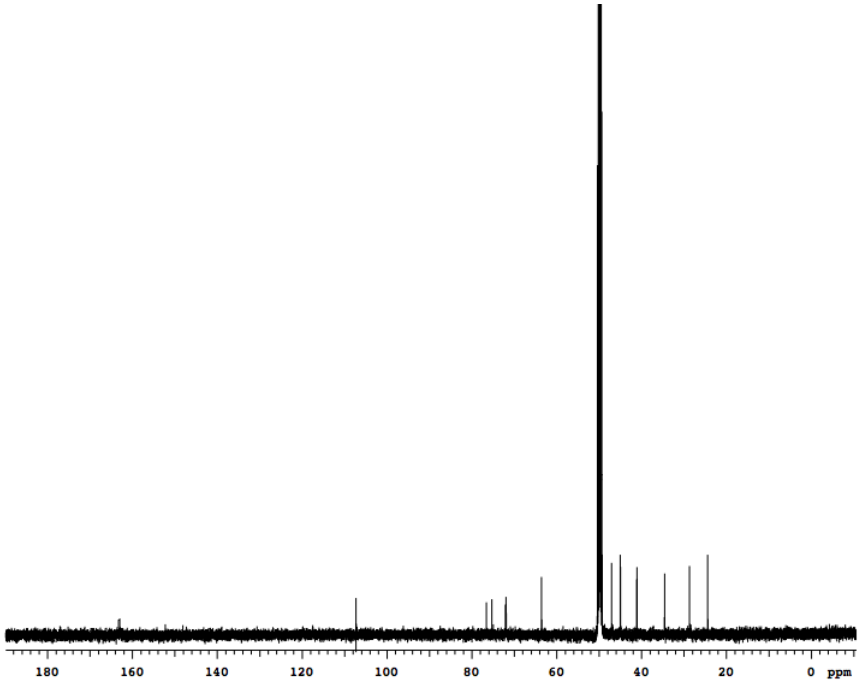
^{13}C NMR (125 MHz, CD_3OD) δ 76.56, 75.25, 72.10, 71.91, 63.56, 47.05, 44.98, 41.10, 34.57, 28.70, 24.40 (107.25 is noise from NMR.)

MS (ESI, m/z) calcd for $C_{11}H_{20}N_3O$ $(M+H)^+$ 210.16, found 210.17

optical rotation $[\alpha]_D^{19.1}$ = +1.4 (*c* = 0.85 in MeOH)



^{1}H NMR of **4f**



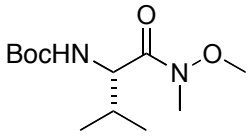
^{13}C NMR of **4f**

G. Syntheses of Pyrrole-based Peptidomimetics 3 and 13

General Procedure and Preparation of Weinreb's amides

The procedure for preparation of Weinreb's amides is the same to one for preparation of compounds **7a** – **h**. The data for Leu and Ile Weinreb's amides were shown in the previous section **7a** and **7c**.

Preparation of Val Weinreb's amide

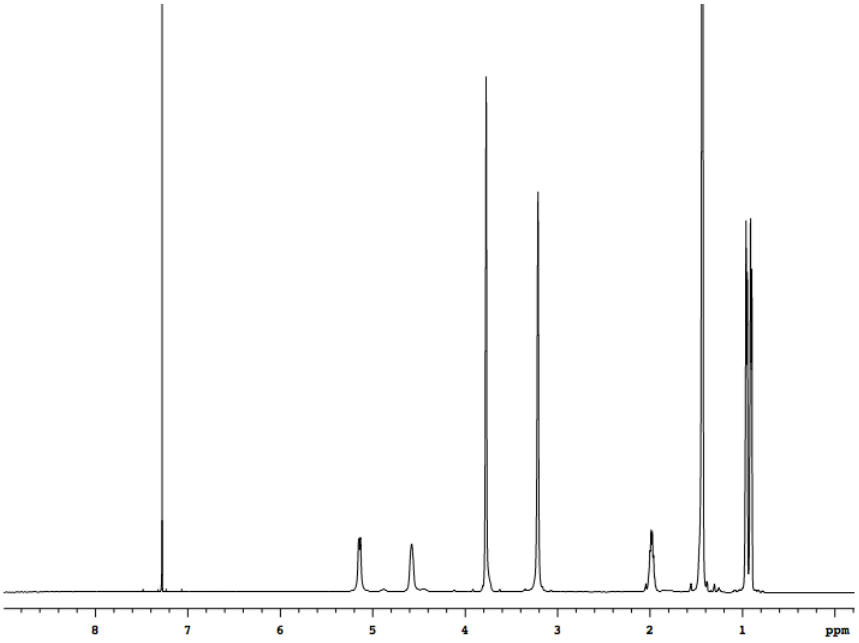


Val Weinreb's amide was prepared from Boc-Val-OH (5.0 g, 23.0 mmol). Flash chromatography (1:5 EtOAc/Hexanes) afforded 5.35 g (89 %) Val Weinreb's amide as a colorless oil.

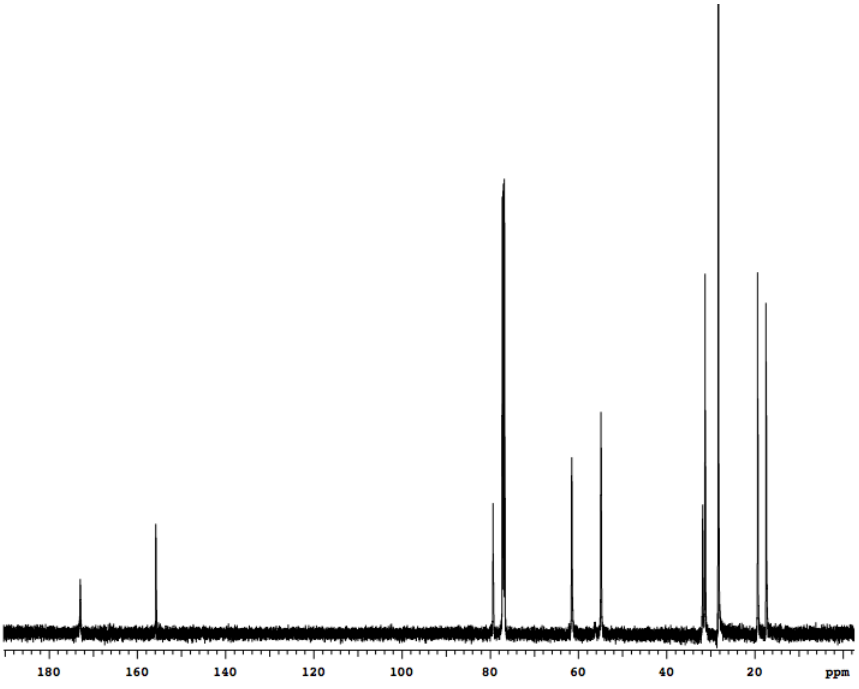
^{1}H NMR (500 MHz, $CDCl_3$) δ 5.14 (d, 1H, J = 9.0 Hz), 4.57 (br, 1H), 3.77 (s, 3H), 3.20 (s, 3H), 1.99-1.96 (m, 1H), 1.42 (s, 9H), 0.95 (d, 3H, J = 6.0 Hz), 0.90 (d, 3H, J = 6.5 Hz)

^{13}C NMR (125 MHz, $CDCl_3$) δ 172.9, 155.8, 79.4, 61.5, 54.9, 31.8, 31.3, 28.3, 19.4, 17.5

MS (ESI, m/z) calcd for $C_{12}H_{25}N_2O_4$ $(M+H)^+$ 261.18, found 261.18

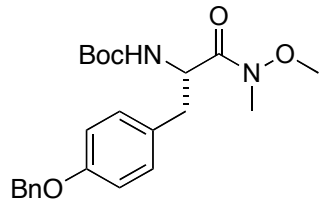


^{1}H NMR of Val Weinreb's amide



^{13}C NMR of Val Weinreb's amide

Preparation of Tyr Weinreb's amide

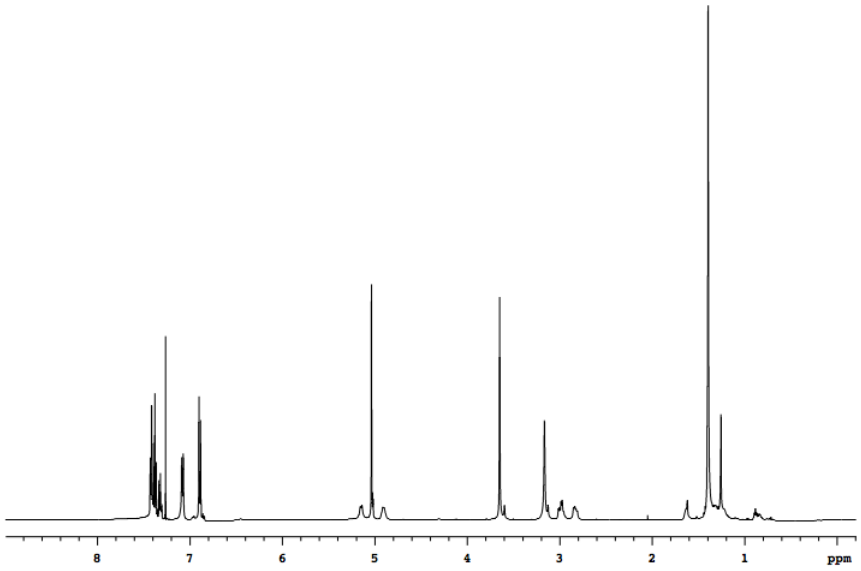


Tyr Weinreb's amide was prepared from Boc-Tyr(Bn)-OH (5.0 g, 13.5 mmol). Flash chromatography (1:4 EtOAc/Hexanes) afforded 5.50 g (99 %) Tyr Weinreb's amide as a white solid.

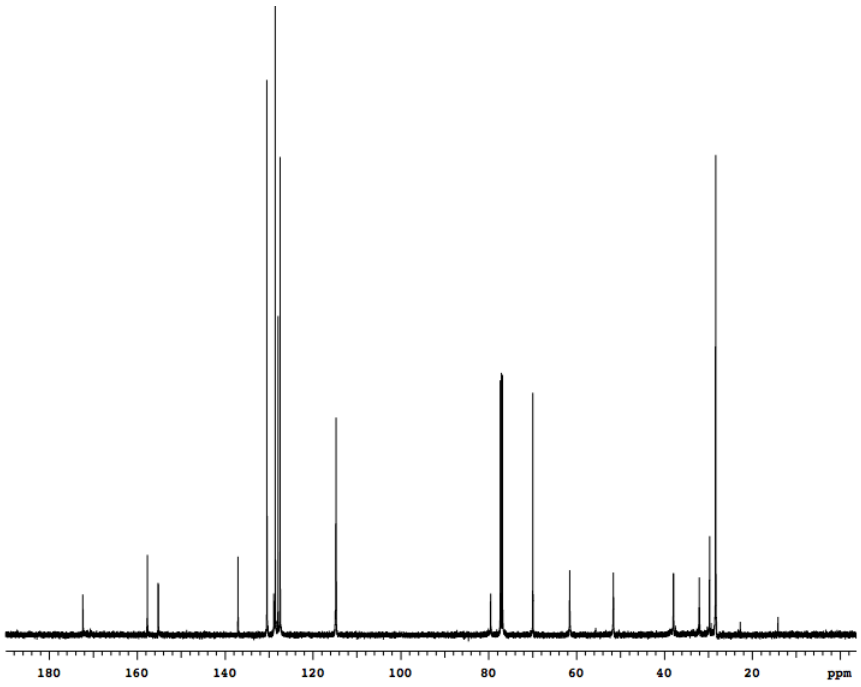
^{1}H NMR (500 MHz, $CDCl_3$) δ7.53 (d, 2H, J = 7.0 Hz), 7.39 (t, 2H, J = 7.5 Hz), 7.33 (t, 1H, J = 7.0 Hz), 7.09 (d, 2H, J = 8.5 Hz), 6.90 (d, 2H, J = 9.0 Hz), 5.16 (d, 1H, J = 8.5 Hz), 5.04 (s, 2H), 4.91 (br, 1H), 3.66 (s, 3H), 3.17 (s, 3H), 3.00 (dd, 1H, J = 6.0, 13.5 Hz), 2.84 (dd, 1H, J = 6.0 Hz, J = 13.0 Hz), 1.40 (s, 9H)

^{13}C NMR (125 MHz, $CDCl_3$) δ 172.3, 157.6, 155.1, 137.0, 130.4, 128.8, 128.5, 127.8, 127.4, 114.6, 79.5, 69.8, 61.5, 51.5, 37.8, 32.0, 28.2

MS (ESI, m/z) calcd for $C_{23}H_{31}N_2O_5$ $(M+H)^+$ 415.22, found 415.23

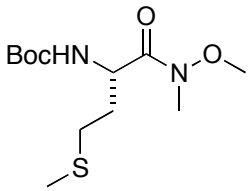


^{1}H NMR of Tyr Weinreb's amide



^{13}C NMR of Tyr Weinreb's amide

Preparation of Met Weinreb's amide

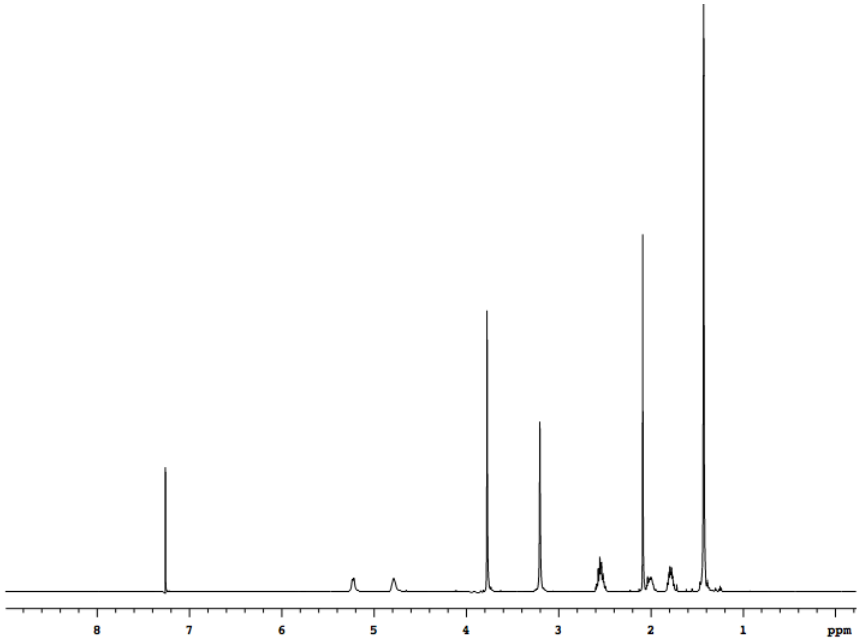


Met Weinreb's amide was prepared from Boc-Met-OH (1.0 g, 4.0 mmol). Flash chromatography (1:3 EtOAc/Hexanes) afforded 1.13 g (80 %) Met Weinreb's amide as a colorless oil.

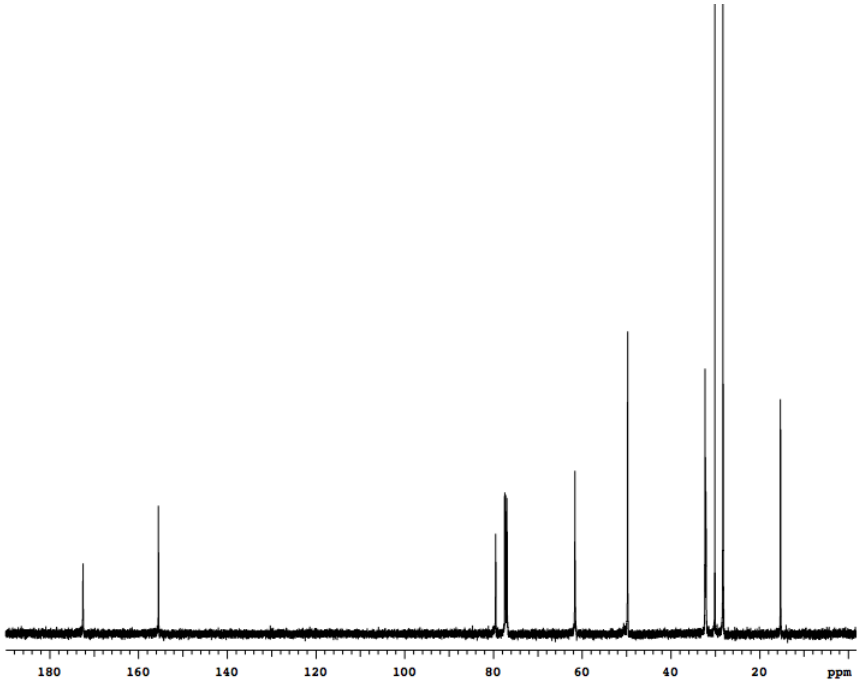
^{1}H NMR (500 MHz, $CDCl_3$) δ 5.24 (d, 1H, J = 8.5 Hz), 4.80 (br, 1H), 3.79(s, 3H), 3.21 (s, 3H), 2.60-2.52 (m, 2H), 2.10 (s, 3H), 1.83-1.73 (m, 2H), 1.44 (s, 9H)

^{13}C NMR (125 MHz, $CDCl_3$) δ 172.3, 155.3, 79.3, 61.4, 49.5, 32.1, 31.9, 29.9, 28.1, 15.1

MS (ESI, m/z) calcd for $C_{12}H_{25}N_2O_4S$ $(M+H)^+$ 293.15, found 293.15

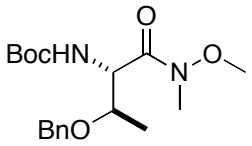


^{1}H NMR of Met Weinreb's amide



^{13}C NMR of Met Weinreb's amide

Preparation of Thr Weinreb's amide



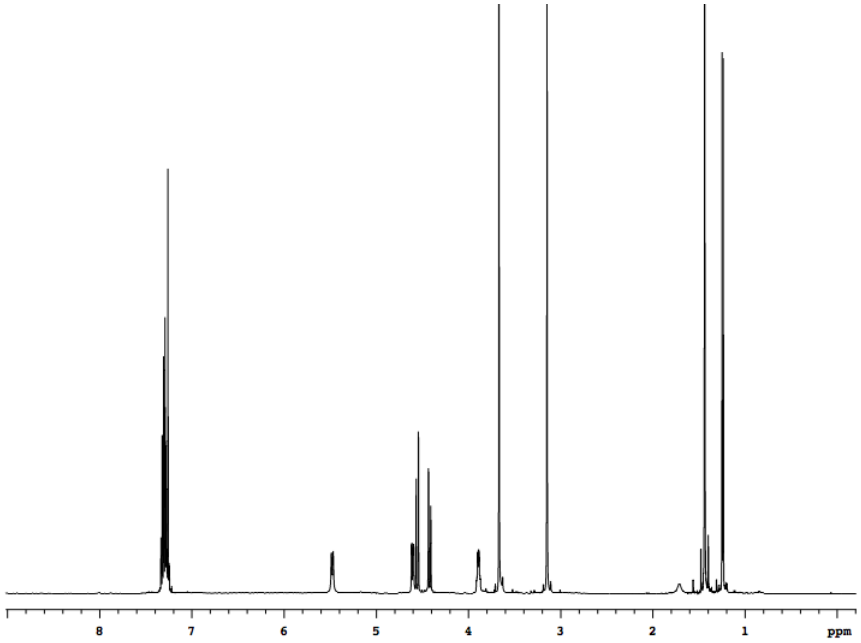
6f

Thr Weinreb's amide was prepared from Boc-Thr(Bn)-OH (4.0 g, 14.0 mmol). Flash chromatography (1:3 EtOAc/Hexanes) afforded 4.30 g (94 %) Thr Weinreb's amide as a colorless oil.

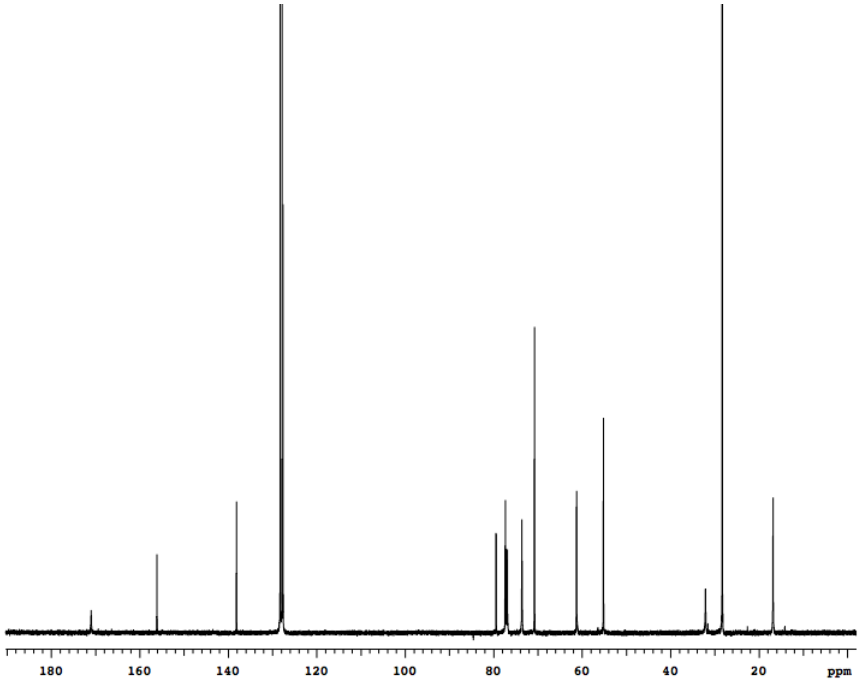
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.35-7.26 (m, 5H), 5.90 (d, 1H, J = 9.0 Hz), 4.62 (dd, 1H, J = 2.5, 9.0 Hz), 4.57 (d, 1H, J = 12.0 Hz), 4.43 (d, 1H, J = 12.5 H), 3.92-3.88 (m, 1H), 3.68 (s, 3H), 3.16 (s, 3H), 1.45 (s, 9H), 1.25 (d, 3H, J = 6.5 Hz)

^{13}C NMR (125 MHz, $CDCl_3$) δ 170.8, 156.0, 138.0, 128.1, 127.7, 127.5, 79.3, 73.5, 70.6, 61.1, 55.0, 32.0, 28.2, 16.7

MS (ESI, m/z) calcd for $C_{18}H_{29}N_2O_5$ $(M+H)^+$ 353.21, found 353.20



^{1}H NMR of Thr Weinreb's amide

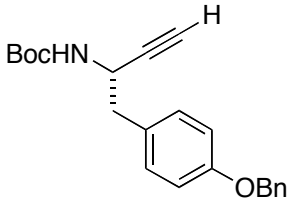


^{13}C NMR of Thr Weinreb's amide

General Procedure and Preparation of Amino-alkyne

The procedure for preparation of amino-alkyne is the same to one for preparation of compounds **8a** – **h**. The data for Leu alkyne were shown in the previous section **8a**.

Preparation of Tyr-alkyne

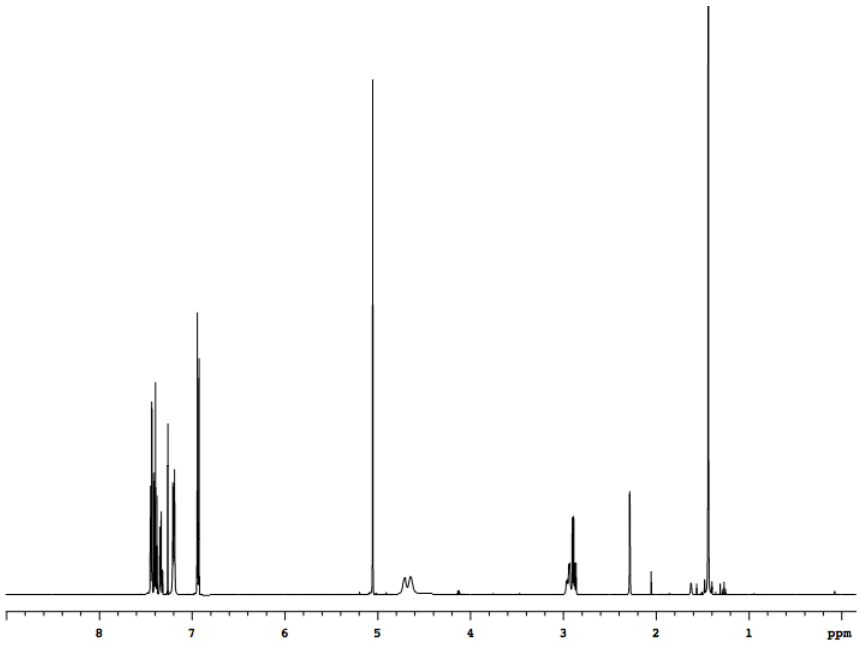


Tyr-alkyne was prepared from Boc-Tyr(Bn)-OH (3.86 g, 10.9 mmol). Flash chromatography (1:5 EtOAc/Hexanes) afforded 3.24 g (85 %) Tyr-alkyne as a white solid.

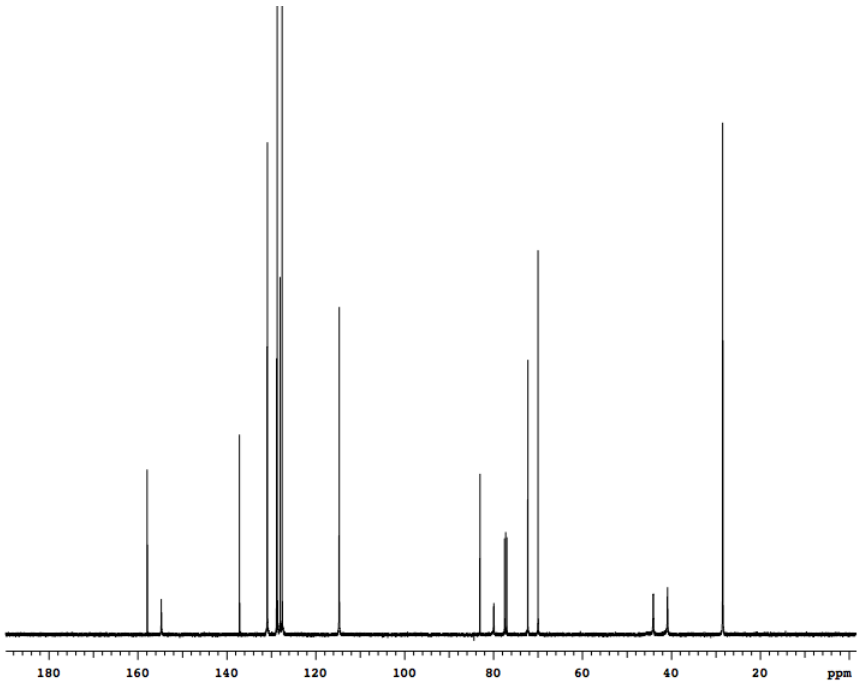
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.45 (d, 2H, J = 7.0 Hz), 7.40 (t, 2H, J = 7.0 Hz), 7.34 (t, 1H, J = 7.0 Hz), 7.20 (d, 2H, J = 3.5Hz), 6.94 (d, 2H, 8.5 Hz), 5.06 (s, 2H), 4.72 (br, 1H), 4.65 (br, 1H),

^{13}C NMR (125 MHz, $CDCl_3$) δ 157.6, 154.5, 136.9, 130.6, 128.5, 128.4, 127.8, 127.3, 114.5, 82.8, 79.7, 72.3, 69.7, 43.8, 40.6, 28.2

MS (ESI, m/z) calcd for $C_{22}H_{26}NO_3$ $(M+H)^+$ 352.19, found 352.17



^{1}H NMR of Tyr-alkyne

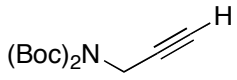


^{13}C NMR of Tyr-alkyne

General Procedure for Gly-alkyne

To a solution of propargylamine (1.0 equiv) in acetonitrie (1.8 M) was added di-tert-butyldicarbonate (2.5 equiv) followed by DMAP (1.0 equiv). The reaction mixture was stirred at 25 ^{o}C for 16 h and then diluted with ethyl acetate. The organic layer was washed with water and brine and dried over $MgSO_4$. After completely removing the solvent, the amino alkyne was purified by flash chromatography.

Preparation of Gly-alkyne



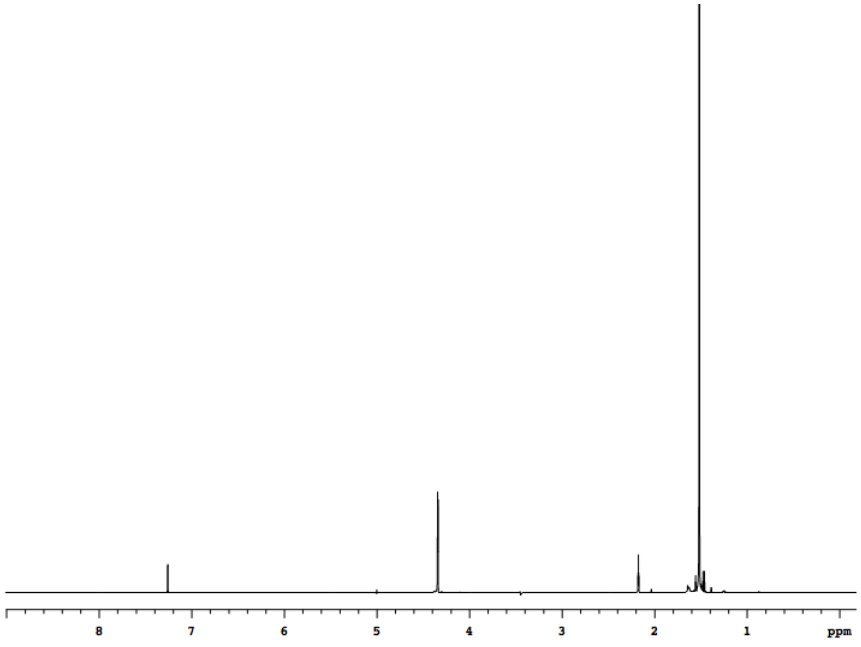
7c

Gly-alkyne was prepared from propagylamine (0.6 g, 10.9 mmol). Flash chromatography (1:19 EtOAc/Hexanes) afforded 0.9 g (32 %) Gly-alkyne as a colorless oil.

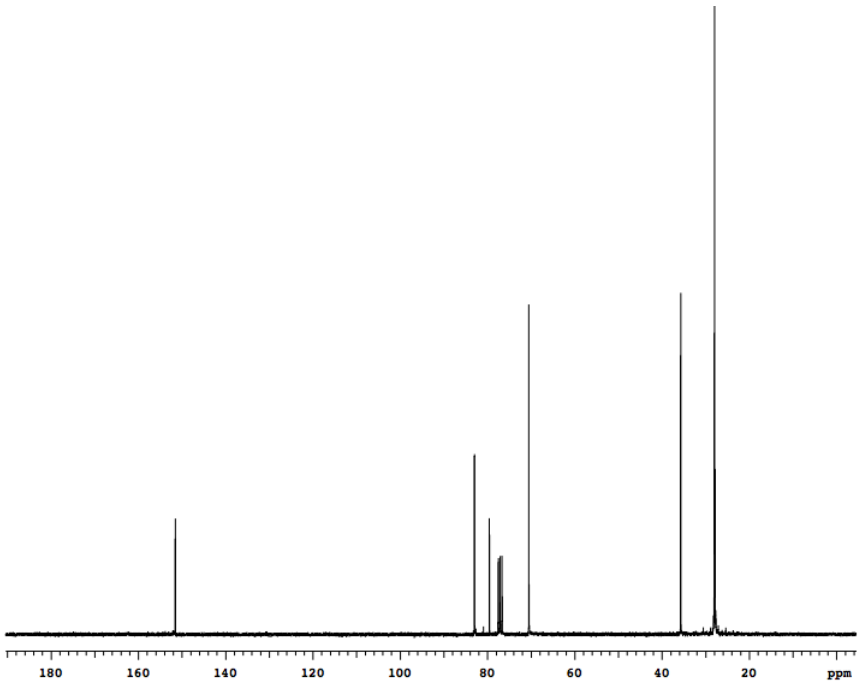
^{1}H NMR (500 MHz, $CDCl_3$) δ 4.35 (d, 2H, J = 4.0 Hz), 2.19 (t, 1H, J = 4.0 Hz), 1.53 (s, 18H)

^{13}C NMR (125 MHz, $CDCl_3$) δ 151.5, 82.9, 79.5, 70.4, 35.6, 27.9

MS (ESI, m/z) calcd for $C_{13}H_{22}NO_4$ $(M+H)^+$ 256.15, found 256.15



^{1}H NMR of Gly-alkyne



^{13}C NMR of Gly-alkyne

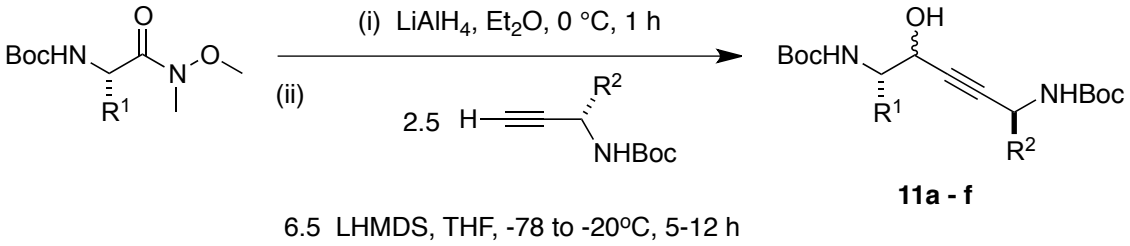
General Procedure and Preparation of Compounds 11a - f

The Weinreb's amide (1.0 equiv) was dissolved in dry ether (0.15 M) and cooled to 0 °C under N_2 (g). Solution of LiAlH (1.0 equiv, 1.0 M in Et_2O) was added via syringe over a period of 20 min. The resulting reaction mixture was stirred at 0 °C for 1 h and then quenched by dropwise addition of 5% $KHSO_4$ (aq.). The phases were separated and the aqueous phase was extracted with ether. The combined organic phases were dried over $MgSO_4$, and then the solvent was removed to get the amino aldehyde, which was used immediately for next step without further purification.

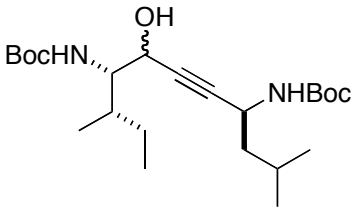
To a solution of the amino alkyne (2.5 equiv) in anhydrous THF (0.3 M) was added lithium bis(trimethylsilyl)amide (6.5 equiv) at -78 ^oC under N_2 (g). The resulting solution was stirred at -78 ^oC for 1 h. The aldehyde (1.0 equiv) in anhydrous THF (0.3 M) was added via cannula over a 20 min period. The reaction mixture was stirred at -78

^{o}C under N_2 (g) for 5 - 12 h. The mixture was then quenched by adding saturated NH_4Cl (aq) and extracted with diethyl ether. The combined ether fractions were dried over $MgSO_4$ and concentrated under vacuum. The compounds **11a - f** were purified by Flash chromatography. [1,2]

Scheme S8. Synthesis of compounds **11a - f**.



Preparation of Compound 11a



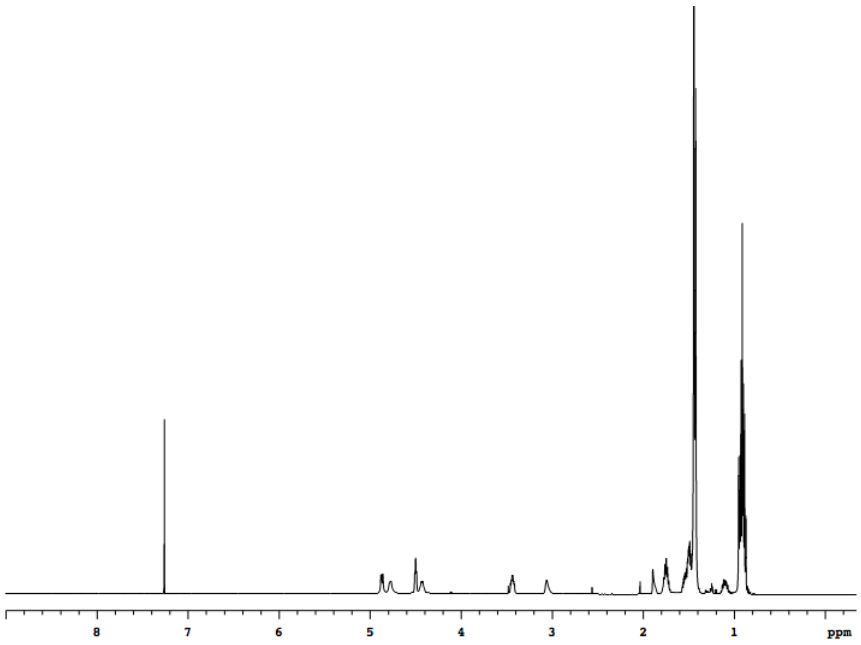
11a

Compound **11a** was prepared from Ile-Weinreb's amide (0.28 g, 1.0 mmol) and Leu-alkyne (0.55 g, 2.5 mmol). Flash chromatography (1:3 EtOAc/Hexanes) afforded 0.25 g (57 %) **11a** as a white solid.

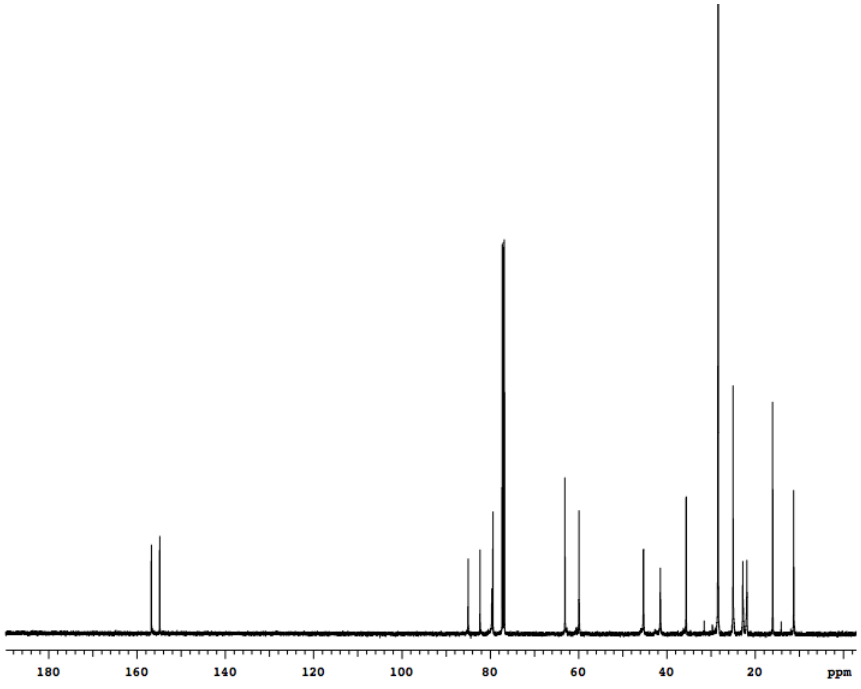
^{1}H NMR (500 MHz, $CDCl_3$) δ 4.88 (d, 1H, J = 9.5 Hz), 4.78 (br, 1H), 4.45-4.43 (m, 1H), 4.61-4.42 (m, 1H), 3.49-3.42 (m, 1H), 3.07 (br, 1H), 1.90 (br, 1H), 1.78-1.73 (m, 2H), 1.58-1.54 (m, 1H), 1.53-1.49 (m, 1H), 1.45 (s, 9H), 1.43 (s, 9H), 0.96-0.88 (m, 12H)

^{13}C NMR (125 MHz, $CDCl_3$) δ 156.7, 154.8, 85.0, 82.3, 79.6, 79.4, 63.0, 59.9, 45.2, 41.4, 35.6, 28.4, 24.9, 22.7, 21.8, 16.0, 11.2

MS (ESI, m/z) calcd for $C_{23}H_{43}N_2O_5$ $(M+H)^+$ 427.32, found 427.30

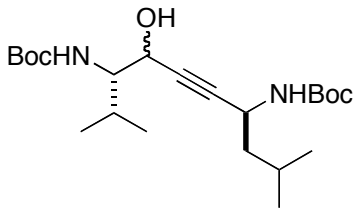


^{1}H NMR of **11a**



^{13}C NMR of **11a**

Preparation of Compound 11b



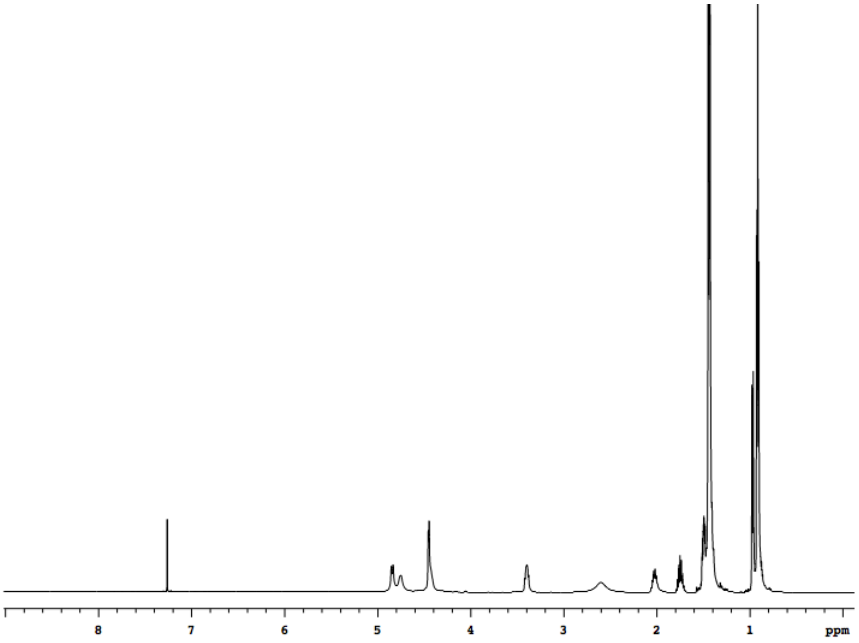
11b

Compound **11b** was prepared from Val-Weinreb's amide (3.70 g, 14.0 mmol) and Leu-alkyne (7.50 g, 35.5 mmol). Flash chromatography (1:5 EtOAc/Hexanes) afforded 1.95 g (33 %) **11b** as a white solid.

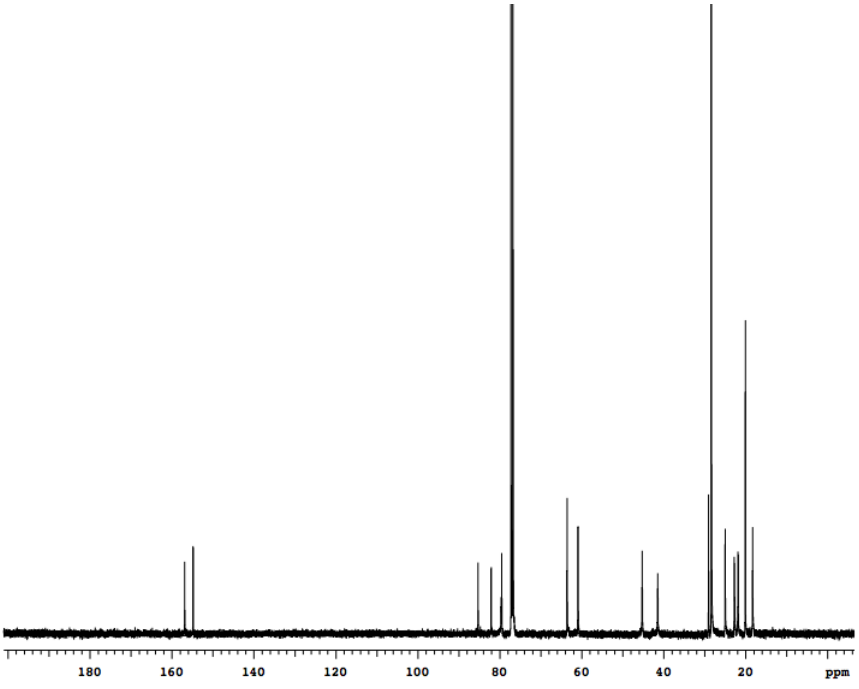
^{1}H NMR (500 MHz, $CDCl_3$) δ 4.85 (d, 1H, J = 9.5 Hz), 4.76 (br, 1H), 4.46 (br, 2H), 3.43-3.39 (m, 1H), 2.61 (br, 1H), 2.05-2.00 (m, 1H), 1.79-1.73 (m, 1H), 1.52-1.49 (m, 2H), 1.46 (s, 9H), 1.44 (s, 9H), 0.98 (d, 3H, J = 9.0 Hz), 0.94-0.92 (m, 9H)

^{13}C NMR (125 MHz, $CDCl_3$) δ 156.9, 154.8, 85.3, 82.1, 79.7, 79.5, 63.6, 60.9, 45.2, 41.5, 29.0, 25.0, 22.8, 21.8, 20.1, 18.3

MS (ESI, m/z) calcd for $C_{22}H_{41}N_2O_5$ $(M+H)^+$ 413.30, found 413.31

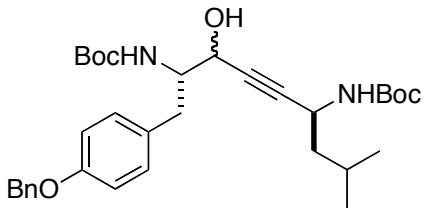


^{1}H NMR of **11b**



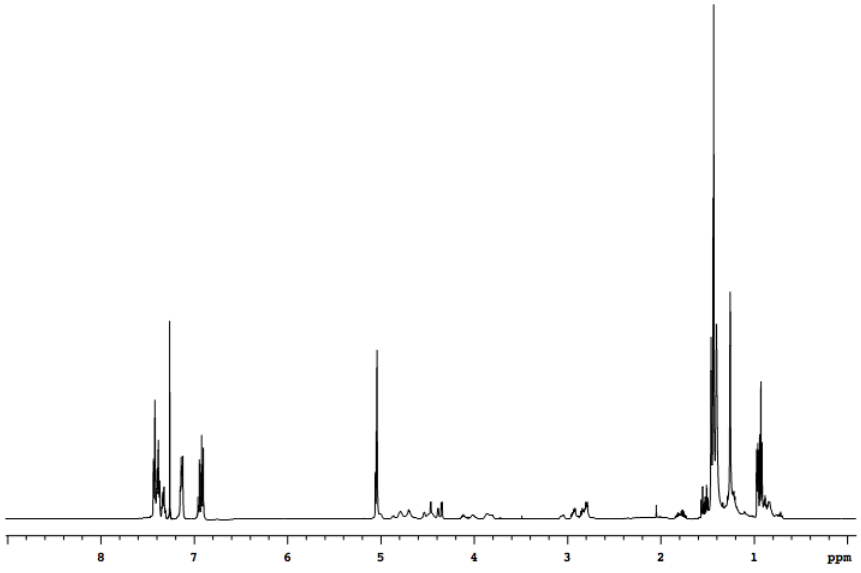
^{13}C NMR of **11b**

Preparation of Compound 11c



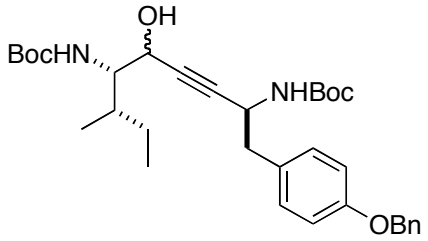
11c

Compound **11c** was prepared from Tyr-Weinreb's amide (0.83 g, 2.0 mmol) and Leu-alkyne (1.06 g, 5.0 mmol). Flash chromatography (1:3 EtOAc/Hexanes) afforded 0.40 g (35 %) **11c** as yellowish solid. Two isomers weren't separated by flash chromatography. MS (MALDI, m/z) calcd for $C_{33}H_{46}N_2NaO_6$ $(M+Na)^+$ 589.33, found 589.44



^{1}H NMR of **11c**

Preparation of Compound 11d



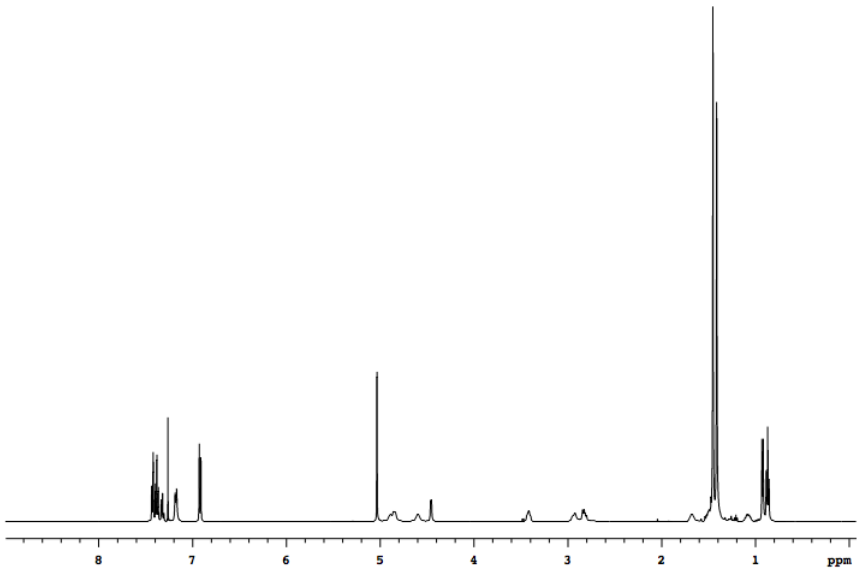
11d

Compound **11d** was prepared from Leu-Weinreb's amide (0.55 g, 2.0 mmol) and Tyr-alkyne (1.02 g, 4.8 mmol). Flash chromatography (1:5 EtOAc/Hexanes) afforded 0.31 g (29 %) **11d** as a yellow oil.

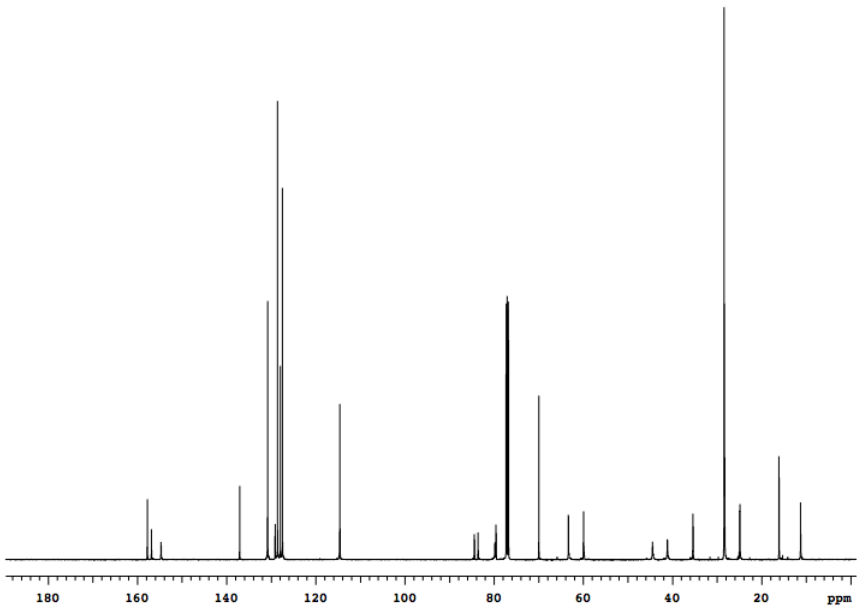
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.43 (d, 2H, J = 7.0 Hz), 7.39 (t, 2H, J = 7.0 Hz), 7.33 (t, 1H, J = 7.0 Hz), 7.19 (d, 2H, J = 8.0 Hz), 6.93 (d, 2H. J = 8.0 Hz), 5.04 (s, 2H), 4.90 (br, 1H), 4.86 (br, 1H), 4.61 (br, 1H), 4.47 (d, 1H, J = 4.0 Hz), 3.43 (br, 1H), 2.93 (br, 1H), 2.83 (dd, 1H, J = 7.5, 12.5 Hz), 1.69 (br, 1H), 1.52 (br, 1H), 1.46 (s, 9H). 1.42 (s, 9H), 1.01-1.08 (m, 1H), 0.93 (d, 3H, J = 6.5 Hz), 0.88 (t, 3H, J = 7.5 Hz)

^{13}C NMR (125 MHz, $CDCl_3$) δ 157.8, 156.9, 154.7, 137.1, 130.8, 130.8, 129.1, 128.6, 127.9, 127.5, 114.6, 84.4, 83.6, 79.8, 79.6, 70.0, 63.3, 59.9, 44.5, 41.1, 35.4, 28.4, 28.3, 24.9, 16.1, 11.3

MS (MALDI, m/z) calcd for $C_{33}H_{47}N_2O_6$ $(M+H)^+$ 589.33, found 589.45

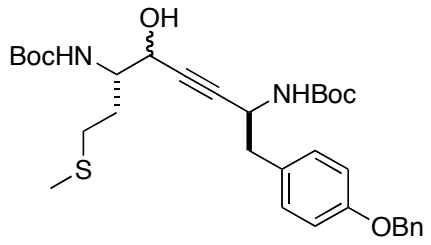


^{1}H NMR of **11d**



^{13}C NMR of **11d**

Preparation of Compound 11e



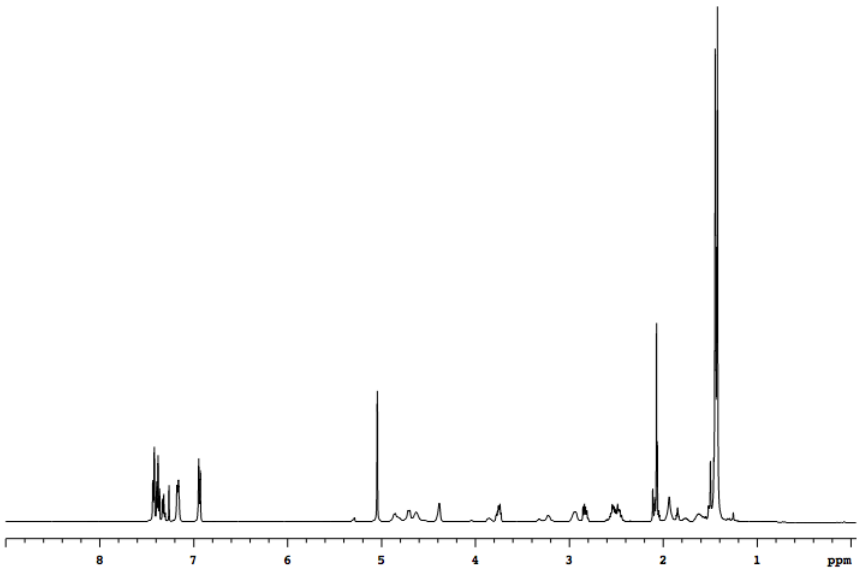
11e

Compound **11e** was prepared from Met-Weinreb's amide (0.29 g, 1.0 mmol) and Tyr-alkyne (0.84 g, 2.4 mmol). Flash chromatography (1:3 to 1:1 EtOAc/Hexanes) afforded 0.17 g (29 %) **11e** as a brown solid.

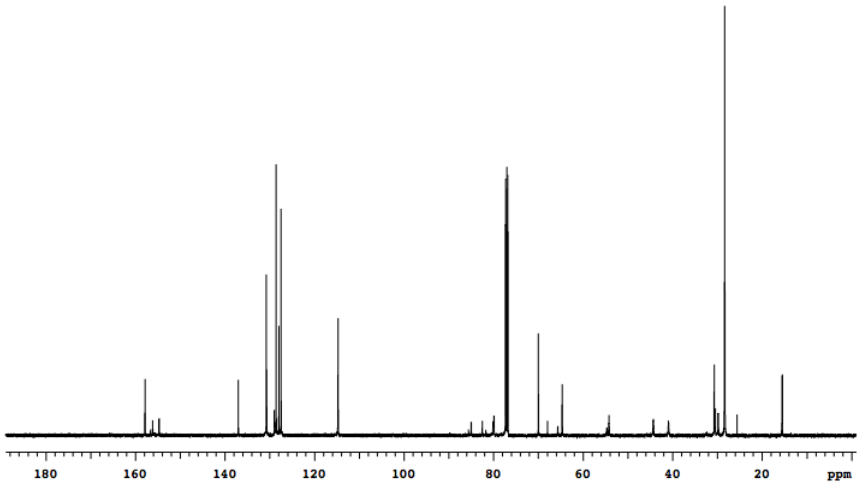
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.42 (d, 2H, J = 7.5 Hz), 7.38 (t, 2H, J = 7.5 Hz), 7.32 (t, 1H, J = 7.5 Hz), 7.17 (d, 2H, J = 8.0 Hz), 6.94 (d, 2H, J = 8.0 Hz), 5.04 (s, 2H), 4.85 (br, 1H), 4.71 (d, 1H, J = 8.5 Hz), 4.63 (br, 1H), 4.38 (br, 1H), 3.78-3.73 (m, 1H), 2.93 (br, 1H), 2.83 (dd, 1H, J = 7.5, 13.5 Hz), 2.57-2.45 (m, 2H), 2.07 (s, 3H), 1.94 (br, 2H), 1.45 (s, 9H), 1.42 (s, 9H)

^{13}C NMR (125 MHz, $CDCl_3$) δ 157.8, 156.1, 154.7, 137.0, 130.7, 128.9, 128.5, 127.9, 127.4, 114.7, 84.9, 82.5, 80.1, 79.8, 69.9, 64.6, 54.2, 44.3, 40.9, 30.7, 28.3, 25.6, 15.5

MS (MALDI, m/z) calcd for $C_{32}H_{44}N_2NaO_6Si$ $(M+Na)^+$ 607.28, found 607.49

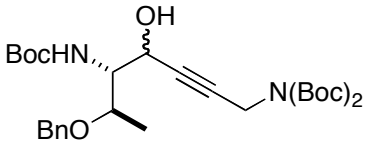


^{1}H NMR of **11e**



^{13}C NMR of **11e**

Preparation of Compound 11f



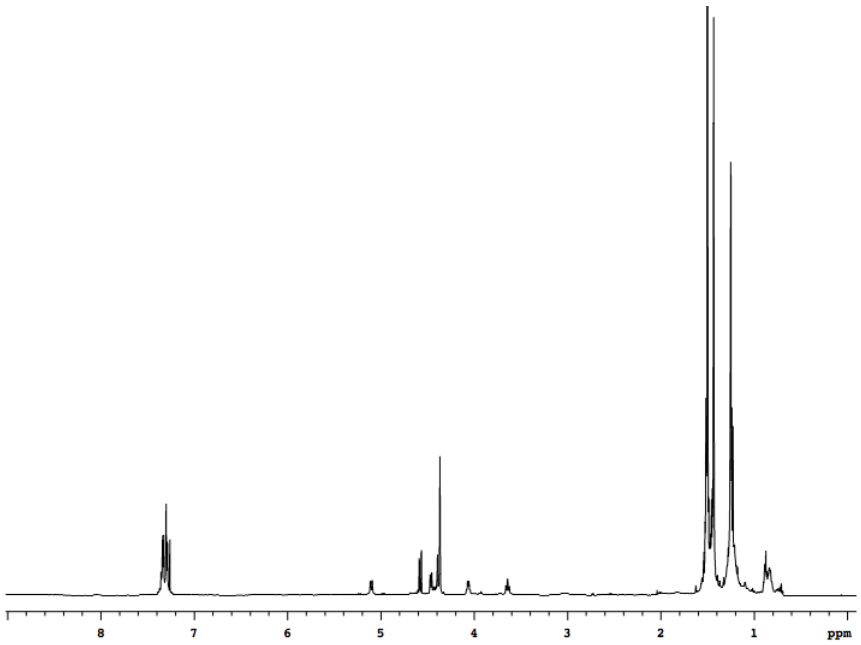
11f

Compound **11f** was prepared from Thr-Weinreb's amide (400.0 mg, 1.14 mmol) and Gly-alkyne (720.0 mg, 2.84 mmol). Flash chromatography (1:5 EtOAc/Hexanes) afforded 74.4 mg (12.0 %) **11f** as a yellowish oil.

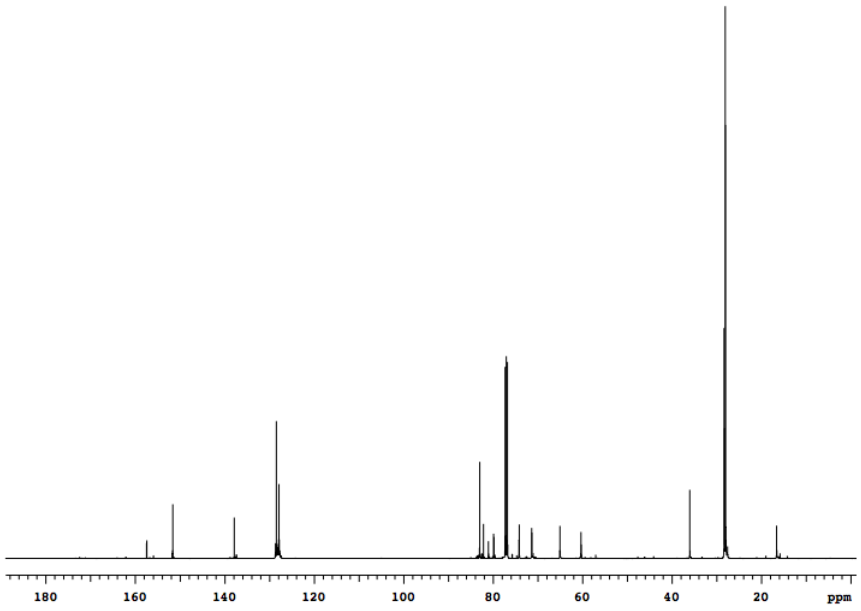
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.41-7.27 (m, 5H), 5.11 (d, 1H, J = 9.5 Hz), 4.58 (d, 1H, J =11.0 Hz), 4.47 (d, 1H, J = 7.5 Hz), 4.40-4.37 (m, 3H), 4.09-4.05 (m, 1H), 3.65 (t, 1H, J = 9.0 Hz), 1.50 (s, 9H), 1.44 (s, 9H), 1.25 (d, J = 5.5 Hz)

^{13}C NMR (125 MHz, $CDCl_3$) δ 157.5, 151.6, 137.9, 128.4, 127.9, 127.9, 83.0, 82.2, 81.2, 79.9, 74.2, 71.3, 65.0, 60.3, 36.0, 28.3, 28.0, 16.6

MS (ESI, m/z) calcd for $C_{29}H_{44}LiN_2O_8$ $(M+Li)^+$ 555.33, found 555.32



^{1}H NMR of **11f**

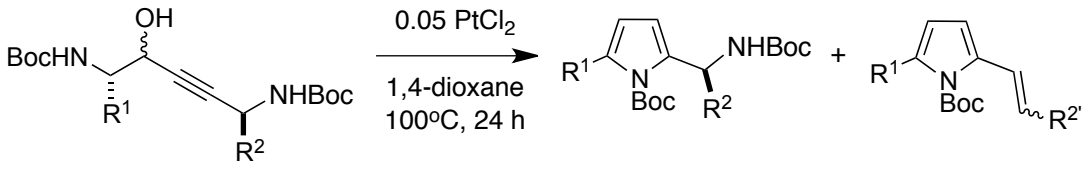


^{13}C NMR of **11f**

General Procedure and Preparation of 12a - f and 14a - f

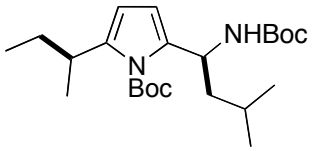
To a solution of the **11a - f** (1.0 equiv) in anhydrous 1,4-dioxane (0.03 M) was added platinum (II) chloride (0.05 equiv). The solution was refluxed for 24 h. The mixture was concentrated under vacuum. The compounds **12a - f** and **14a - f** were purified by Flash chromatography.

Scheme S9. Synthesis of compounds **12a - f** and **14a - f**.



11a - f **12a - f** **14a - f**

Preparation of Compound 12a



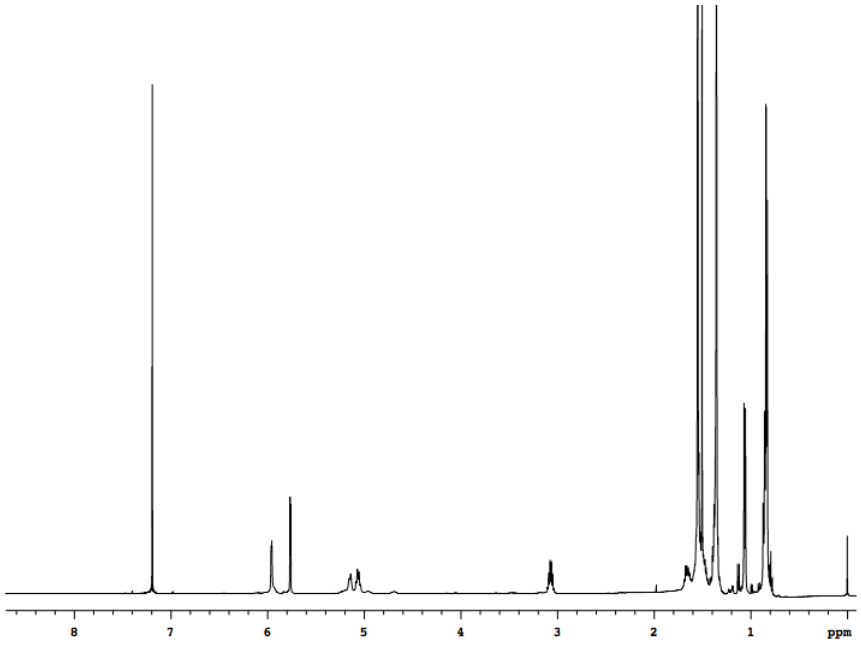
12a

Compound **12a** was prepared from **11a** (97.8 mg, 0.23 mmol). Flash chromatography (3:97 EtOAc/Hexanes) afforded 76.7 mg (82 %) **12a** as a yellowish oil.

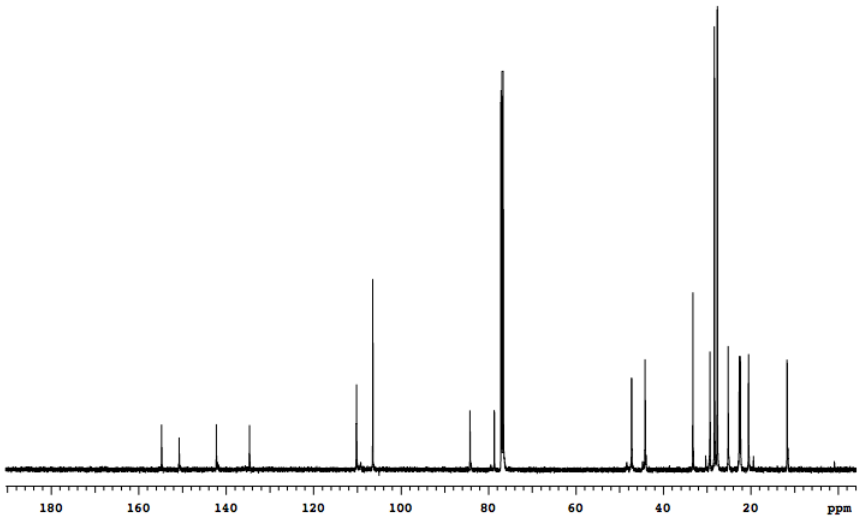
^{1}H NMR (500 MHz, $CDCl_3$) δ 6.04 (d, 1H, J = 3.5 Hz), 5.84 (dd, 1H, J = 0.5, 3.5 Hz), 5.22 (d, 1H, J = 8.0 Hz), 5.14 (dt, 1H, J = 7.0, 15.0 Hz), 3.17-3.13(m, 1H), 1.76-1.71 (m, 1H), 1.63 (s, 9H), 1.60-1.53 (m, 4H), 1.14 (d, 3H, J = 6.5 Hz), 0.99-0.91 (m, 9H)

^{13}C NMR (125 MHz, $CDCl_3$) δ 154.9, 150.9, 142.3, 134.8, 110.3, 106.6, 84.3, 78.8, 47.4, 44.3, 33.4, 29.4, 28.4, 27.8, 25.3, 22.7, 22.5, 20.6, 11.8

MS (MALDI, m/z) calcd for $C_{23}H_{40}N_2NaO_4$ $(M+Na)^+$ 431.29, found 431.37

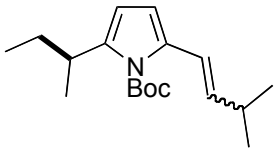


^{1}H NMR of **12a**



^{13}C NMR of **12a**

Preparation of Compound 14a



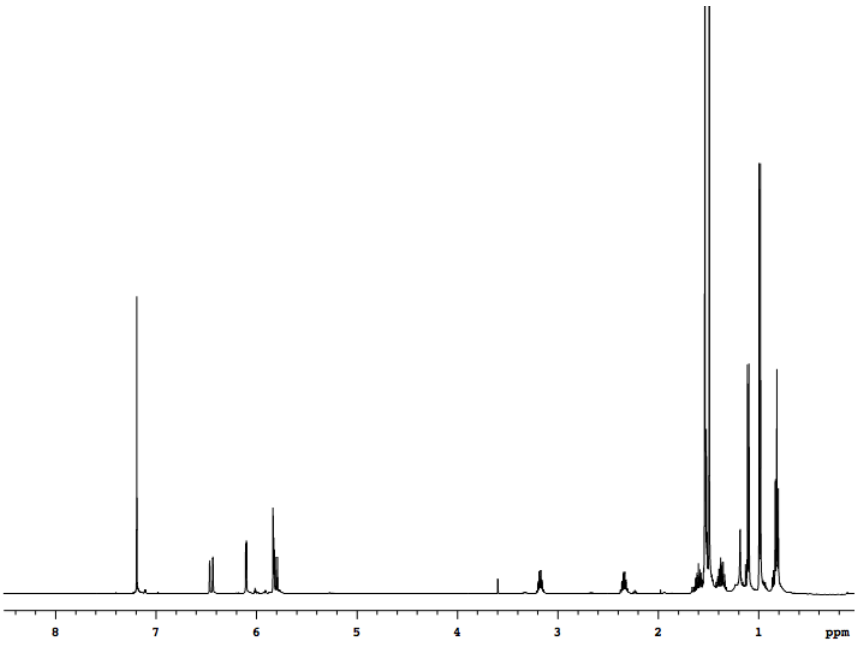
14a

Compound **14a** was prepared from **11a** (97.8 mg, 0.23 mmol). Flash chromatography (3:97 EtOAc/Hexanes) afforded 8.3 mg (12 %) **14a** as a yellow oil.

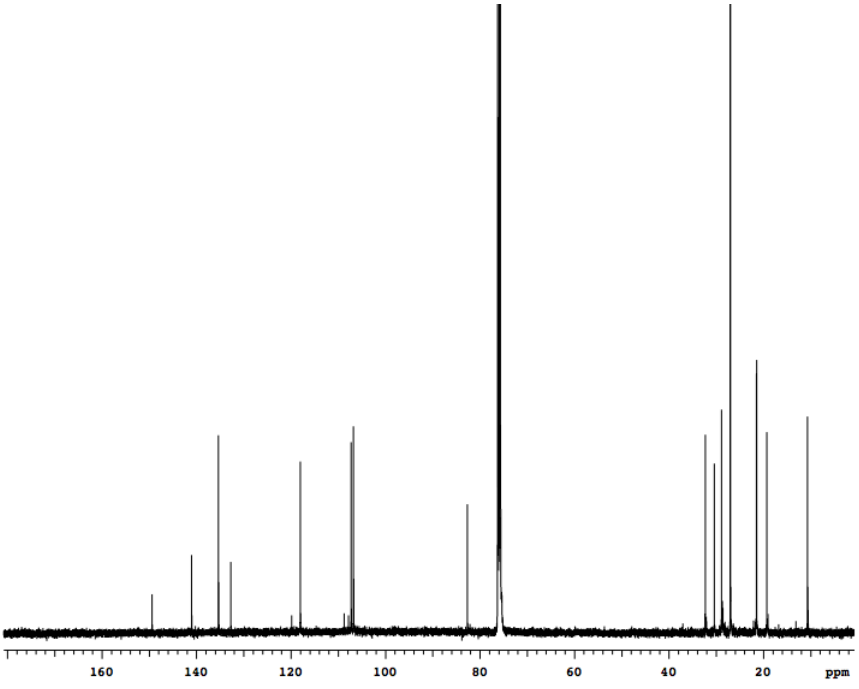
^{1}H NMR (500 MHz, $CDCl_3$) δ 6.53 (d, 1H, J = 16.0 Hz), 6.18 (d, 1H, J = 3.5 Hz), 5.91 (d, 1H, J = 4.0 Hz), 5.89 (dd, 1H, J = 7.0, 16.0 Hz), 3.29-3.22 (m, 1H), 2.45-2.38 (m, 1H), 1.61 (s, 9H), 1.72-1.64 (m, 1H), 1.50-1.42 (m, 1H), 1.18 (d, 3H, J = 7.0 Hz), 1.18 (d, 6H, J = 6.5 Hz), 0.90 (t, 3H, J = 7.5 Hz)

^{13}C NMR (125 MHz, $CDCl_3$) δ 150.4, 142.0, 136.3, 133.7, 119.0, 108.3, 107.8, 83.7, 33.3, 31.4, 29.9, 28.0, 22.5, 22.5, 20.3, 11.7

MS (ESI, m/z) calcd for $C_{18}H_{30}NO_2$ $(M+H)^+$ 292.23, found 292.22

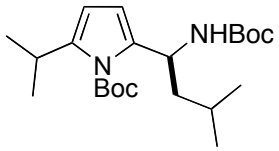


^{1}H NMR of **14a**



^{13}C NMR of **14a**

Preparation of Compound 12b



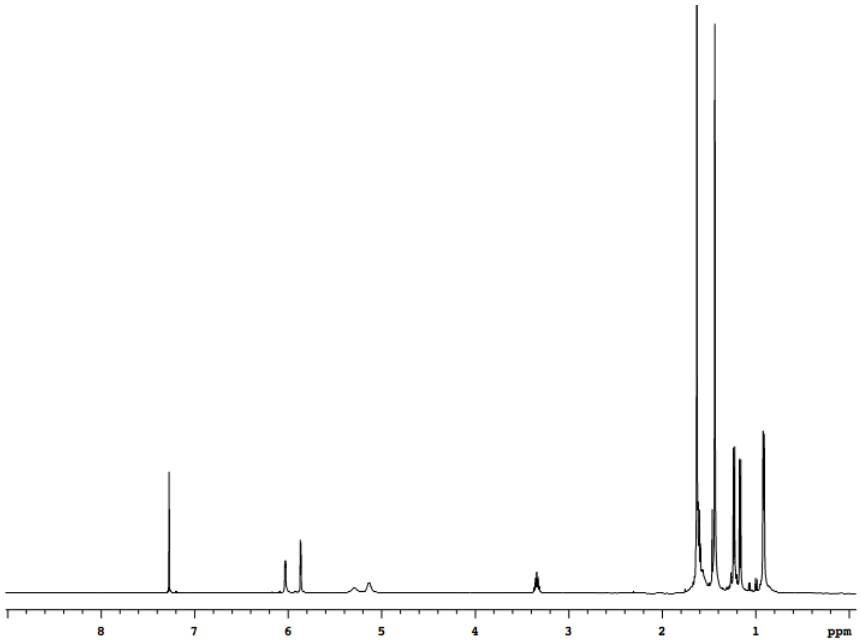
12b

Compound **12b** was prepared from **11b** (297.3 mg, 0.72 mmol). Flash chromatography (3:97 EtOAc/Hexanes) afforded 204.2 mg (72 %) **12b** as a yellowish oil.

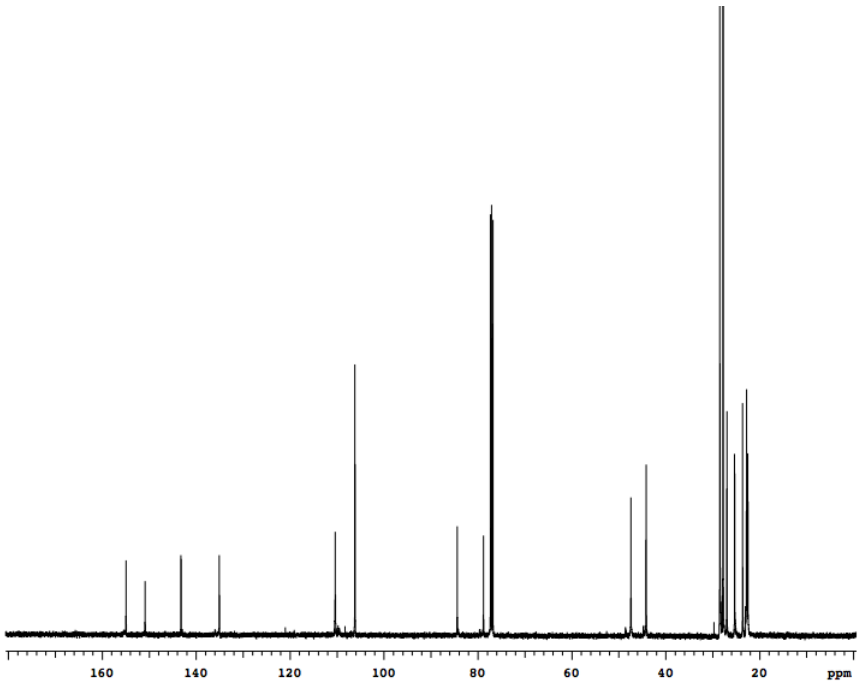
^{1}H NMR (500 MHz, $CDCl_3$) δ 6.03 (d, 1H, J = 3.0 Hz), 5.86 (dd, 1H. J = 1.0, 3.5 Hz), 5.29 (br, 1H), 5.13 (br, 1H), 3.37-3.31 (m, 1H), 1.67-1.48 (m, 12H), 1.44 (s, 9H), 1.23 (d, 3H, J = 7.0 Hz), 1.16 (d, 3H, J = 6.5 Hz), 0.91(d, 6H, J = 6.5 Hz)

^{13}C NMR (125 MHz, $CDCl_3$) δ 154.9, 150.8, 143.2, 135.0, 110.3, 106.1, 84.3, 78.8, 47.4, 44.1, 28.4, 27.7, 26.9, 25.3, 23.6, 22.7, 22.7, 22.5

MS (ESI, m/z) calcd for $C_{22}H_{39}N_2O_4$ $(M+H)^+$ 395.29, found 395.29

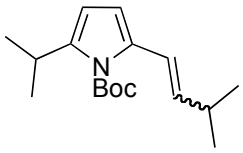


^{1}H NMR of **12b**



^{13}C NMR of **12b**

Preparation of Compound 14b



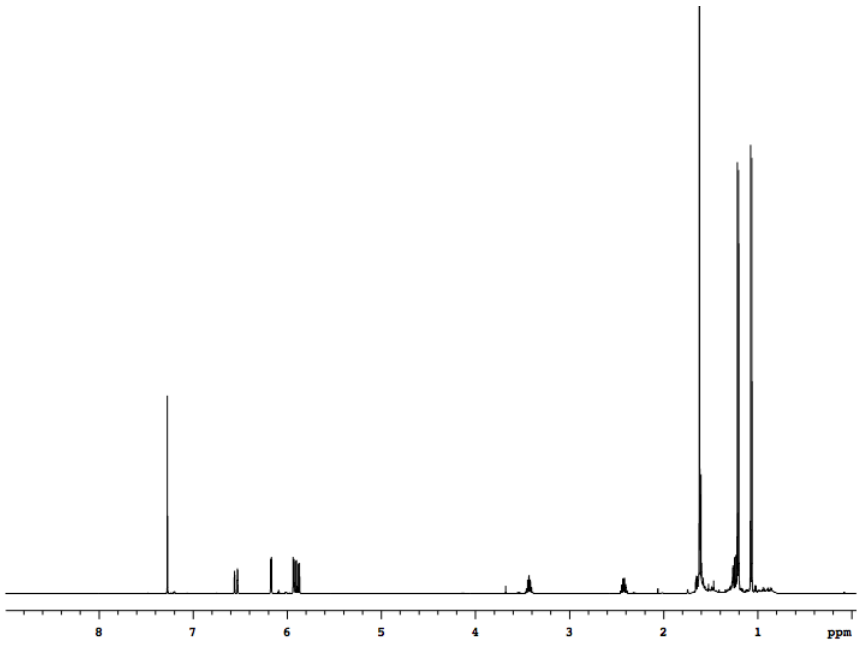
14b

Compound **14b** was prepared from **11b** (297.3 mg, 0.72 mmol). Flash chromatography (3:97 EtOAc/Hexanes) afforded 5.7 mg (3 %) **14b** as a yellowish oil.

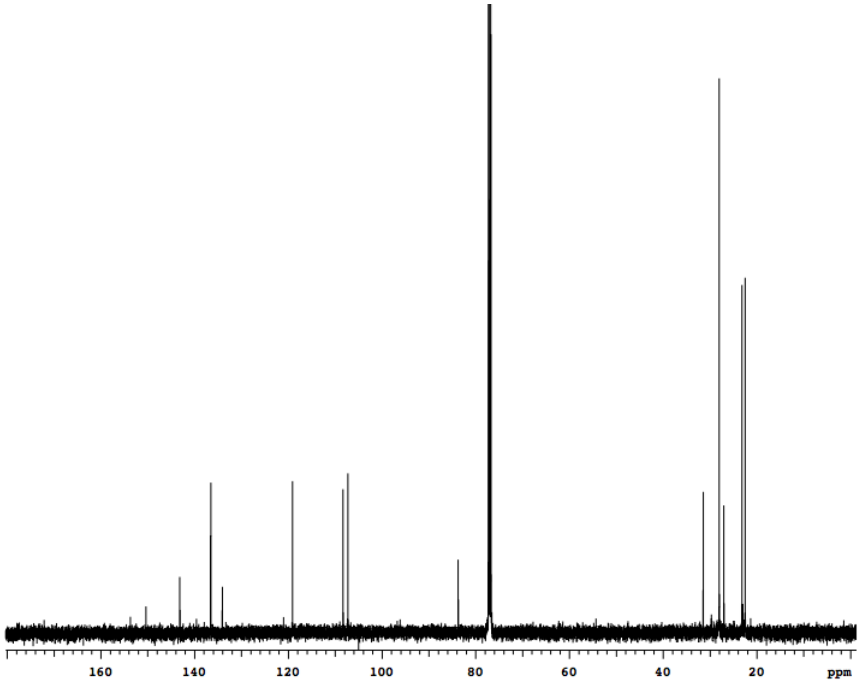
^{1}H NMR (500 MHz, $CDCl_3$) δ 6.54 (d, 1H, J = 16 Hz), 6.17 (d, 1H, J = 3.5 Hz), 5.93 (d, 1H, J = 3.5 Hz), 5.89 (dd, 1H, J = 7.0, 16.0 Hz), 3.45-3.40 (m, 1H), 2.42 (m, 1H), 1.62 (s, 9H), 1.21 (d, 6H, J = 6.5 Hz), 1.06 (d, 6H, J = 7.0 Hz)

^{13}C NMR (125 MHz, $CDCl_3$) δ 150.3, 143.1, 136.5, 134.1, 119.1, 108.3, 107.2, 83.7, 31.4, 28.0, 27.0, 23.2, 23.0, 22.5

MS (ESI, m/z) calcd for $C_{17}H_{28}NO_2$ $(M+H)^+$ 278.21, found 278.21

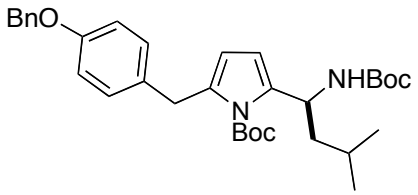


^{1}H NMR of **14b**



^{13}C NMR of **14b**

Preparation of Compound 12c



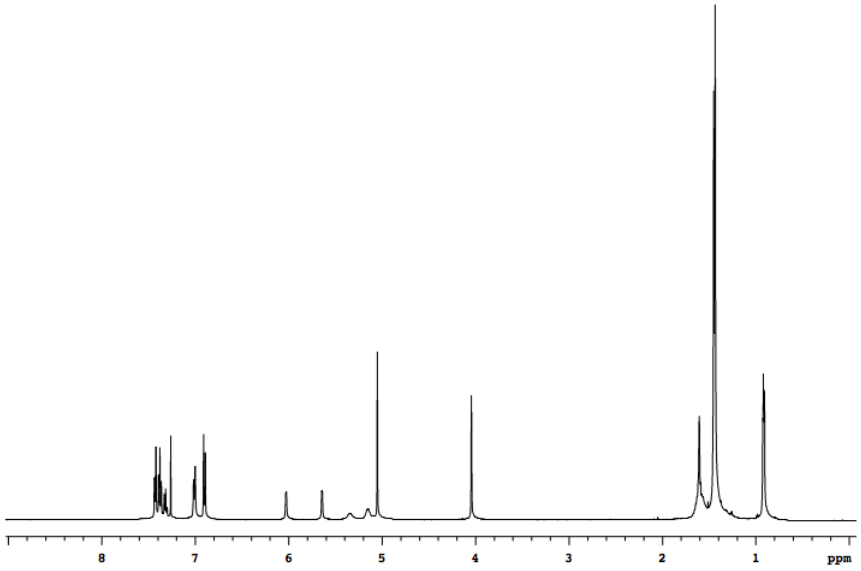
12c

Compound **12c** was prepared from **11c** (93.0 mg, 0.16 mmol). Flash chromatography (1:9 EtOAc/Hexanes) afforded 19.2 mg (21 %) **12c** as a yellowish oil.

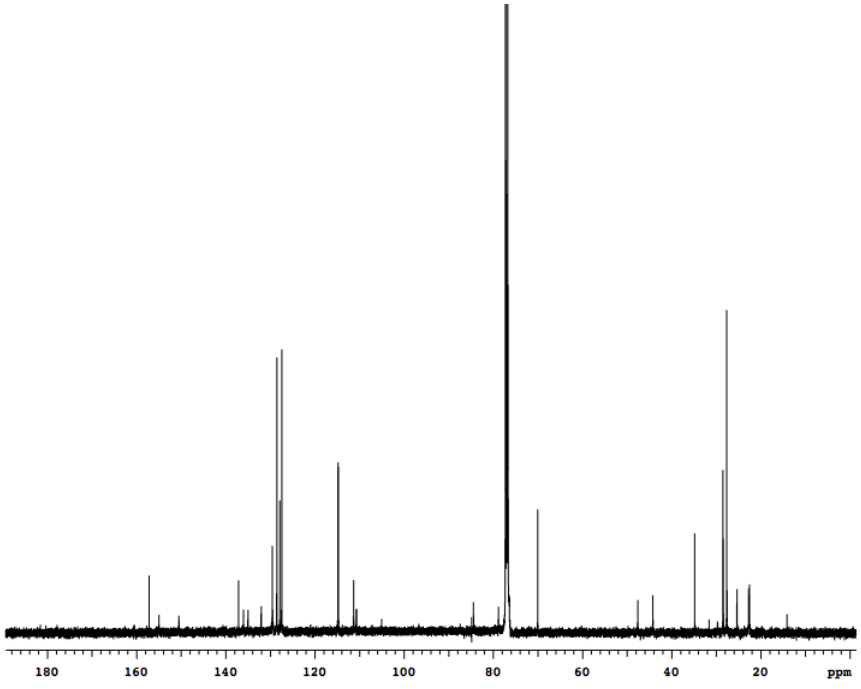
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.44 (d, 2H, J = 8.0 Hz), 7.39 (t, 2H, J = 7.5 Hz), 7.32 (t, 1H, J = 8.0 Hz), 7.02 (d, 2H, J = 8.0 Hz), 6.91 (d, 2H, 8.5 Hz), 6.03 (d, 1H, J = 2.5 Hz), 5.65 (d, 1H, J = 2.5 Hz), 5.35 (br, 1H), 5.16 (br, 1H), 5.06 (s, 2H), 4.05 (s, 2H), 1.61 (br, 3H), 1.46 (s, 9H), 1.44 (s, 9H), 0.93 (d, 3H, 3.0 Hz), 0.92 (d, 3H, 3.0 Hz)

^{13}C NMR (125 MHz, $CDCl_3$) δ 157.2, 155.0, 150.5, 137.1, 136.0, 135.0, 132.0, 129.6, 128.5, 127.9, 127.4, 114.8, 111.3, 110.7, 105.0, 84.4, 78.8, 70.0, 47.6, 44.2, 34.8, 28.5, 27.6, 25.3, 22.7, 22.6

MS (ESI, m/z) calcd for $C_{33}H_{44}LiN_2O_5$ $(M+Li)^+$ 555.34, found 555.33

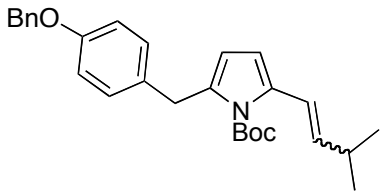


^{1}H NMR of **12c**



^{13}C NMR of **12c**

Preparation of Compound 14c



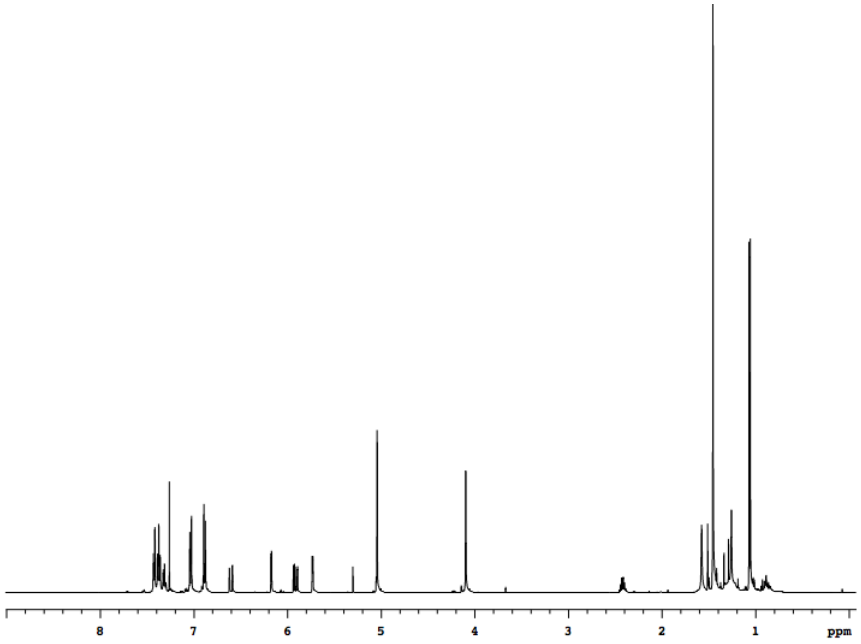
14c

Compound **14c** was prepared from **11c** (93.0 mg, 0.16 mmol). Flash chromatography (1:9 EtOAc/Hexanes) afforded 7.2 mg (10 %) **14c** as a yellowish oil.

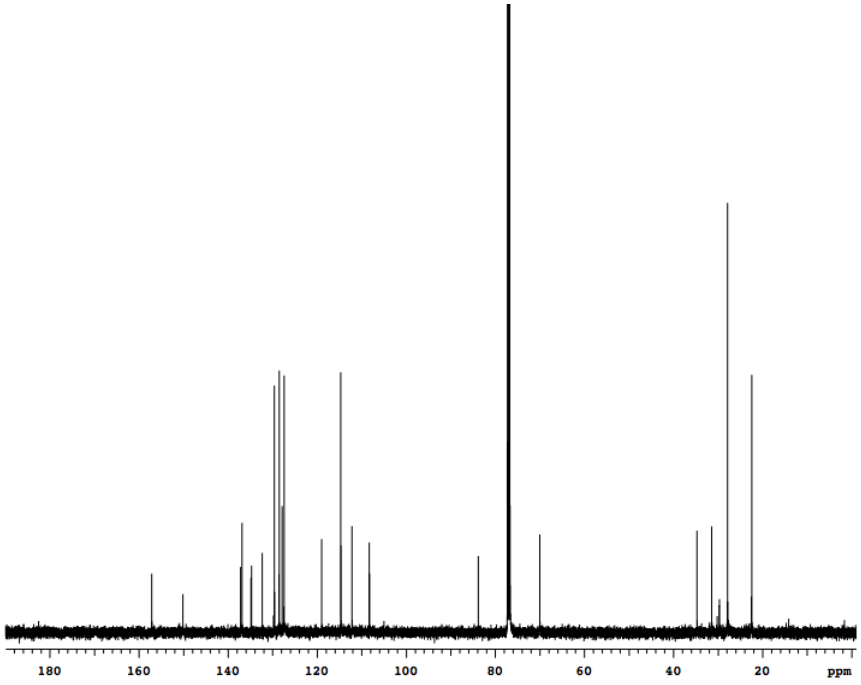
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.43 (d, 2H, J = 7.0 Hz), 7.38 (t, 2H, J = 9.0 Hz), 7.32 (t, 1H, J = 7.0 Hz), 7.04 (d, 2H, J = 9.0 Hz), 6.90 (d, 2H, J = 9.0 Hz), 6.61 (d, 1H, J = 16.0 Hz), 6.18 (d, 1H, J = 3.5 Hz), 5.92 (dd, 1H, J = 6.5, 15.5 Hz), 5.74 (d, 1H, J = 3.5), 5.05 (s, 2H), 4.10 (s, 2H), 2.47-2.40 (m, 1H), 1.46 (s, 9H), 1.07 (d, 6H, J = 6.5 Hz)

^{13}C NMR (125 MHz, $CDCl_3$) δ 157.1, 150.1, 137.2, 136.9, 134.8, 134.7, 132.3, 129.6, 128.5, 127.9, 127.4, 119.0, 114.7, 112.2, 108.3, 83.8, 70.0, 34.7, 31.4, 27.9, 22.4

MS (ESI, m/z) calcd for $C_{28}H_{34}NO_3$ $(M+H)^+$ 432.25, found 432.23

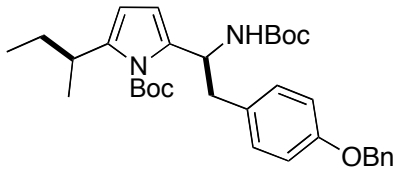


^{1}H NMR of **14c**



^{13}C NMR of **14c**

Preparation of Compound 12d



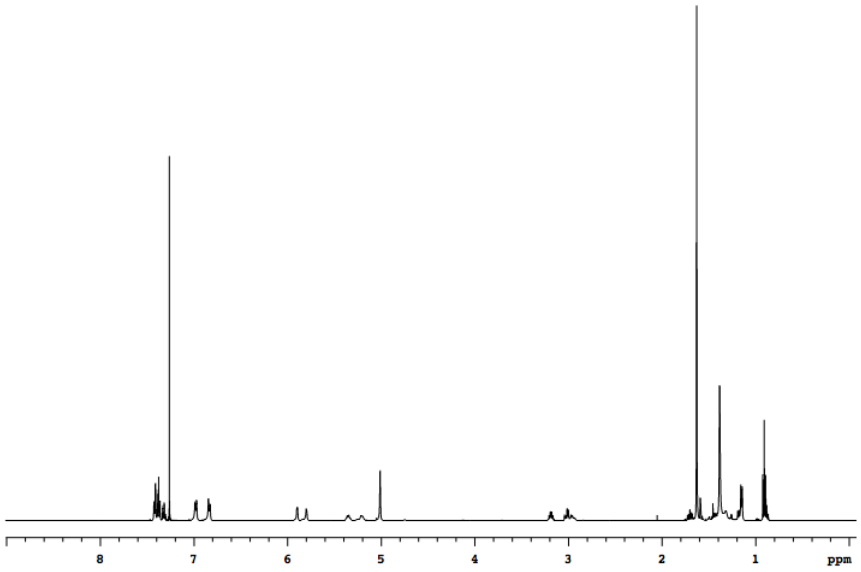
12d

Compound **12d** was prepared from **11d** (164.8 mg, 0.29 mmol). Flash chromatography (1:5 EtOAc/Hexanes) afforded 98.0 mg (61 %) **12d** as a yellowish oil.

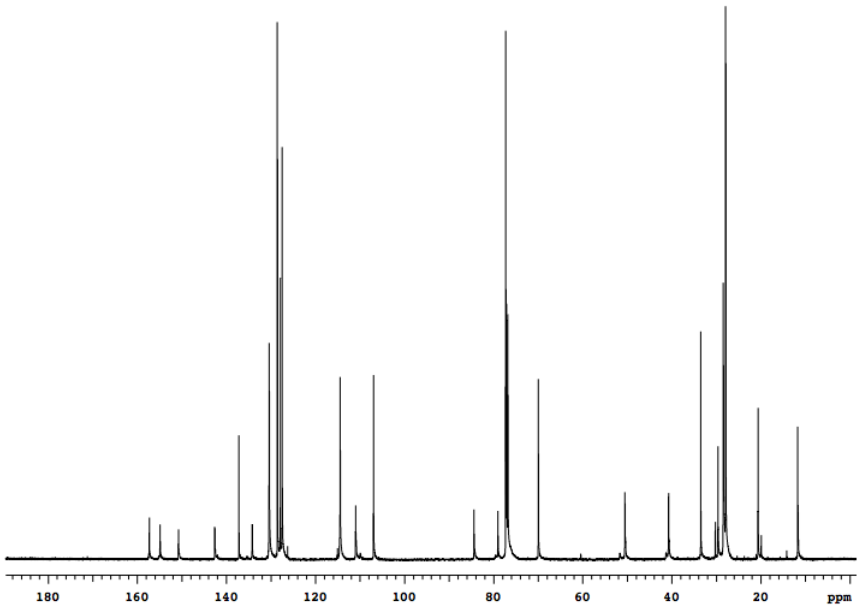
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.42 (d, 2H, J = 7.0 Hz), 7.38 (t, 2H, J = 7.5 Hz), 7.33 (t, 1H, J = 7.0 Hz), 6.99 (d, 2H, J = 8.0 Hz), 6.84 (d, 2H, J = 8.5 Hz), 5.90 (d, 1H, J = 2.5 Hz), 5.81 (d, 1H, J = 3.0 Hz), 5.37-5.34 (td, 1H, J = 6.5, 14.5 Hz), 5.21 (d, 1H, J = 8.5 Hz), 5.02 (s, 2H), 3.23-3.16 (m, 1H), 3.03 (dd, 1H, J = 7.0, 14.0 Hz), 2.96 (dd, 1H, J = 6.5, 13.5 Hz), 1.74-1.67 (m, 1H), 1.64 (s, 9H), 1.47-1.42 (m, 1H), 1.39 (s, 9H), 1.16 (d, 3H, J = 7.0 Hz), 0.92 (t, 3H, J = 7.0 Hz)

^{13}C NMR (125 MHz, $CDCl_3$) δ 157.3, 154.9, 150.9, 142.6, 137.2, 134.2, 130.4, 130.3, 128.6, 127.9, 127.5, 114.5, 111.0, 106.9, 84.4, 79.0, 70.0, 50.5, 40.7, 33.5, 29.6, 28.4, 27.9, 20.6,11.7

MS (MALDI, m/z) calcd for $C_{33}H_{44}N_2NaO_5$ $(M+Na)^+$ 571.31, found 571.43

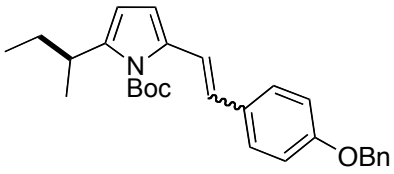


^{1}H NMR of **12d**



^{13}C NMR of **12d**

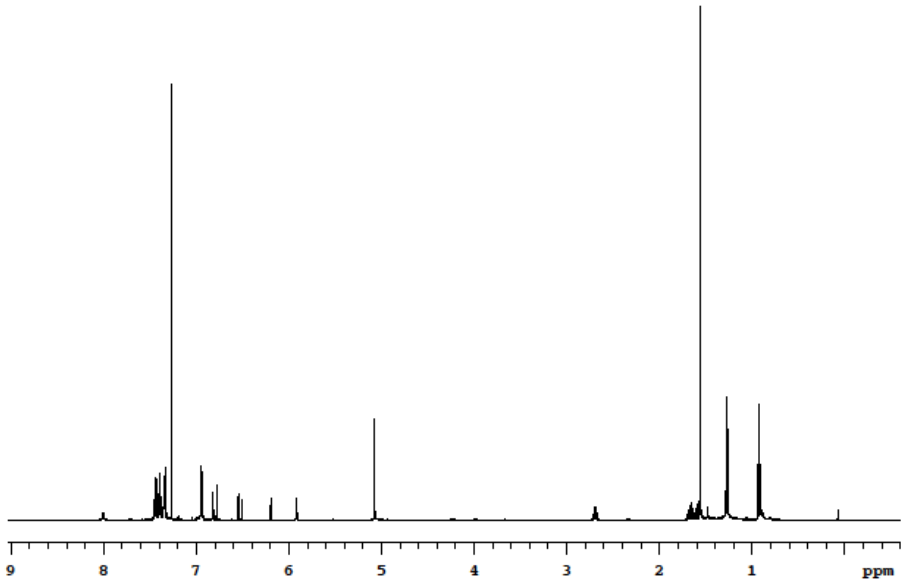
Preparation of Compound 14d



14d

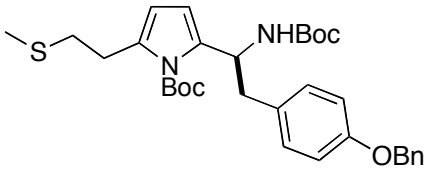
Compound **14d** was prepared from **11d** (164.8 mg, 0.29 mmol). Flash chromatography (1:5 EtOAc/Hexanes) afforded 21.0 mg (17 %) **14d** as a yellowish solid. Most products were a mixture of isomers, and they were not separated easily by flash chromatography. ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.01 (s, 1H), 7.42 (d, 2H, J = 7.0 Hz), 7.40 (t, 2H, J = 7.5 Hz), 7.35 (d, 2H, J = 8.5 Hz), 6.95 (d, 2H, J = 8.5 Hz), 6.80 (d, 1H, J = 16.5 Hz), 6.53 (d, 1H, J = 16.5 Hz), 6.20 (t, 1H, J = 3.0 Hz), 5.93 (t, 1H, J = 3.0 Hz), 5.08 (s, 2H), 2.74-2.7 (m, 1H), 1.72-1.66 (m, 1H), 1.64-1.60 (m, 1H), 1.57 (s, 9H), 1.28 (d, 3H, J = 3.0 Hz), 0.93 (t, 3H, J = 7.5 Hz)

MS (MALDI, m/z) calcd for $C_{28}H_{34}NO_3$ $(M+H)^+$ 432.25, found 432.27



^{1}H NMR of **14d**

Preparation of Compound 12e



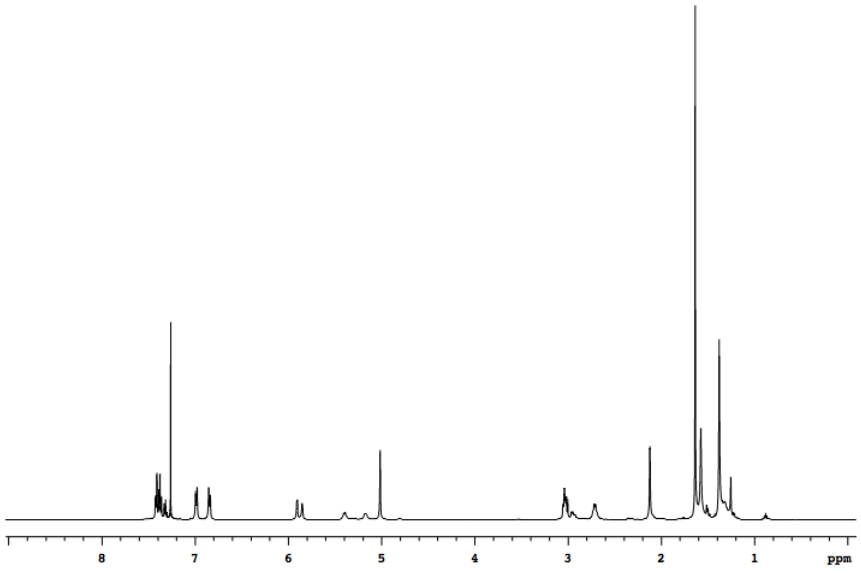
12e

Compound **12e** was prepared from **11e** (30.0 mg, 0.052 mmol). Flash chromatography (3:97 EtOAc/Hexanes) afforded 8.8 mg (30.0 %) **12e** as a yellowish solid.

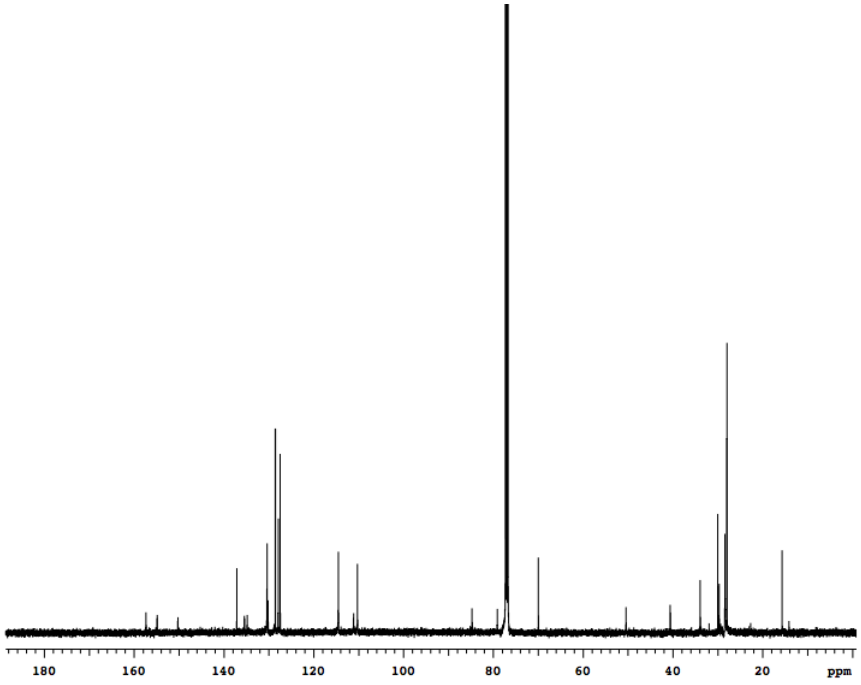
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.42 (d, 2H, J = 7.0 Hz), 7.38 (t, 2H, J = 7.0 Hz), 7.32 (t, 1H, J = 7.0 Hz), 6.99(d, 2H, J = 8.5 Hz), 6.85 (d, 2H, J = 8.5 Hz), 5.91 (d, 1H, J = 3.0 Hz), 5.85 (d, 1H, J = 3.0Hz), 5.43-5.41 (m, 1H), 5.17 (d, 1H, J = 8.0 Hz), 5.02 (s, 2H), 3.06-3.01 (m, 2H), 2.74-2.79 (m, 3H), 2.95 (dd, 1H, J = 6.0 Hz), J = 13.5 Hz), 2.13 (s, 3H), 1.64 (s, 9H), 1.38 (s, 9H)

^{13}C NMR (125 MHz, $CDCl_3$) δ 157.3, 150.9, 150.3, 137.1, 135.4, 134.8, 130.4, 130.3, 128.5, 127.9, 127.5, 114.5, 111.1, 110.3, 84.7, 79.1, 70.0, 50.4, 40.6, 33.9, 30.0, 28.4, 28.0, 15.7

MS (MALDI, m/z) calcd for $C_{32}H_{42}N_2NaO_5S$ $(M+Na)^+$ 589.27, found 589.45

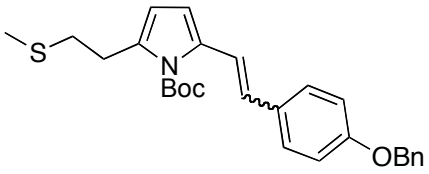


^{1}H NMR of **12e**



^{13}C NMR of **12e**

Preparation of Compound 14e



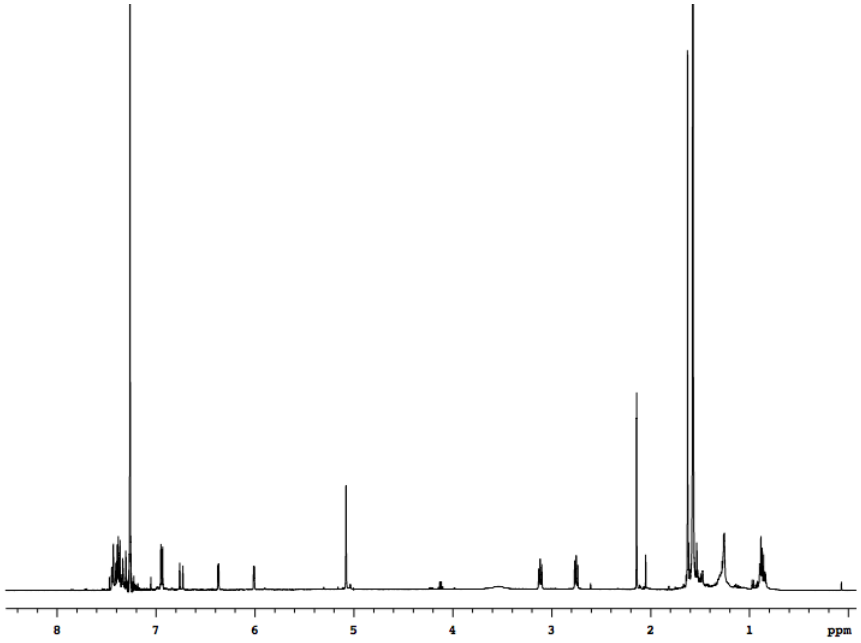
14e

Compound **14e** was prepared from **11e** (30.0 mg, 0.052 mmol). Flash chromatography (3:97 EtOAc/Hexanes) afforded trace amount as a yellowish solid.

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.47-7.30 (m, 9H), 6.94 (d, 2H, J = 8.5 Hz), 6.74 (d, 1H. J = 16.0 Hz), 6.37 (d, 1H, J = 3.5 Hz), 6.01 (d, 1H, J = 3.5 Hz), 5.08 (s, 2H), 3.12 (t, 2H, J = 7.5 Hz), 2.75 (t, 2H, J = 7.5 Hz), 2.14 (s, 3H), 1.57 (s, 9H)

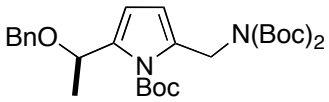
^{13}C NMR was not obtained.

MS (MALDI, m/z) calcd for $C_{27}H_{32}NO_3S$ $(M+H)^+$ 450.21, found 450.20



^{1}H NMR of **14e**

Preparation of Compound 12f



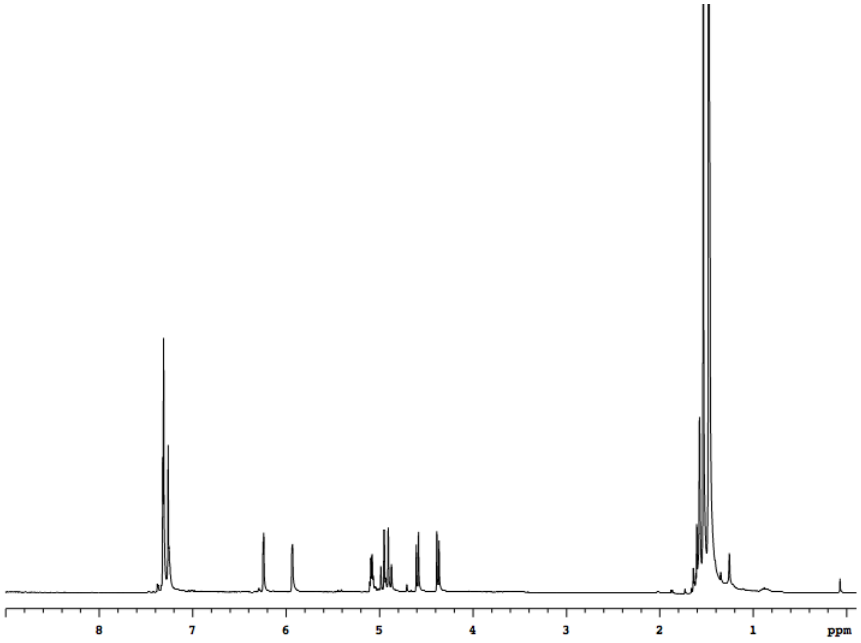
12f

Compound **12f** was prepared from **11f** (47.9 mg, 0.087 mmol). Flash chromatography (1:5 EtOAc/Hexanes) afforded 6.2 mg (13 %) **12f** as a yellow oil.

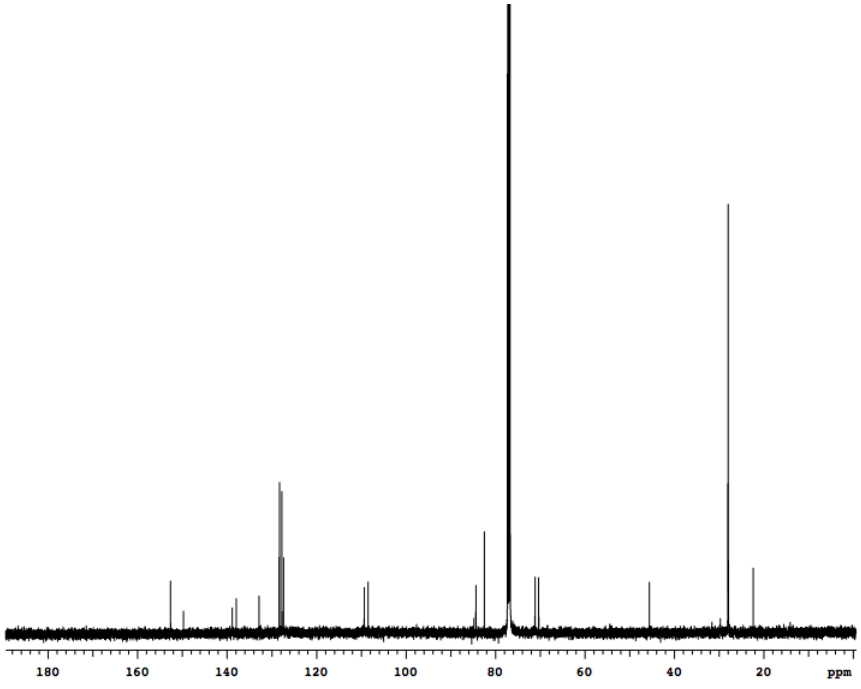
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.32-7.25 (m, 5H), 6.24 (d, 1H, J = 3.5 Hz), 5.93 (d, 1H, J = 3.5 Hz), 5.09 (q, 1H, J = 6.0, 12.5 Hz), 4.97 (d, 1H, J = 17.5 Hz), 4.89 (d, 1H, J = 17.0 Hz), 4.60 (d, 1H, J = 12.0 Hz), 4.38 (d, 1H, J = 12.0 Hz), 1.58 (br, 3H), 1.54 (s, 9H), 1.48 (s, 18H)

^{13}C NMR (125 MHz, $CDCl_3$) δ 152.6, 149.7, 138.8, 137.9, 132.8, 128.3, 127.7, 127.4, 109.3, 108.5, 84.4, 82.5, 71.1, 70.3, 45.6, 28.0, 27.9, 22.4

MS (MALDI, m/z) calcd for $C_{29}H_{42}N_2NaO_7$ $(M+Na)^+$ 553.29, found 553.35

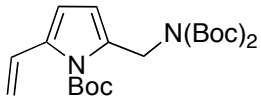


^{1}H NMR of **12f**



^{13}C NMR of **12f**

Preparation of Compound 14f



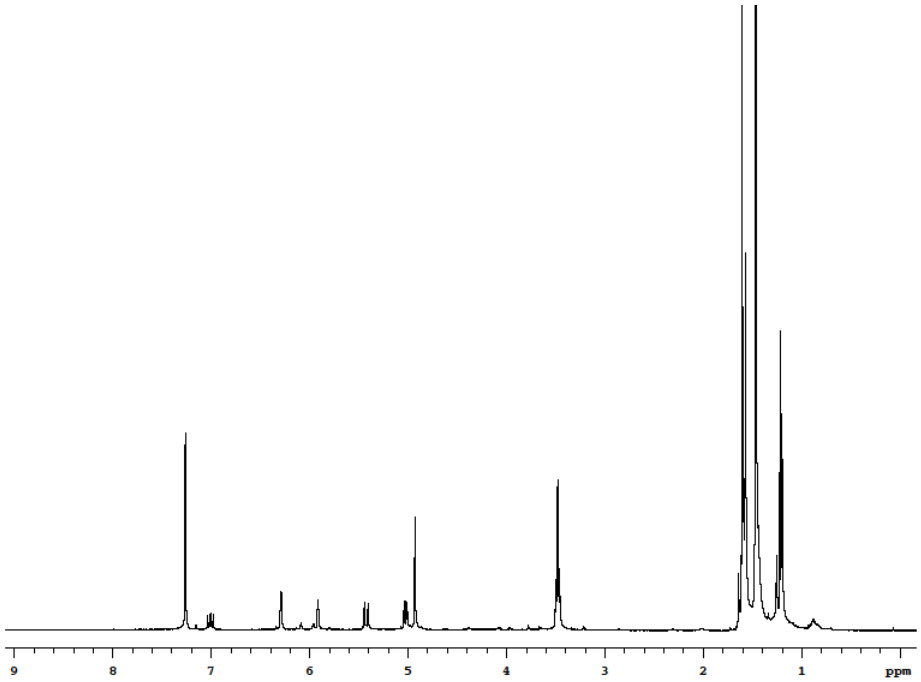
14f

Compound **14f** was prepared from **11f** (47.9 mg, 0.087 mmol). Flash chromatography (1:5 EtOAc/Hexanes) afforded trace amount.

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.02 (dd, 1H, J = 11.5, 17.5 Hz), 6.30 (d, 1H, J = 3.5 Hz), 5.92 (d, 1H, J = 3.0 Hz), 5.43 (d, 1, J = 17.5 Hz), 5.04 (d, 1H, J = 11.0 Hz), 4.94 (s, 2H), 1.61 (s, 9H), 1.48 (s, 18H) (3.48 and 1.22 from diethyl ether contamination.)

^{13}C NMR was not obtained.

MS (MALDI, m/z) calcd for $C_{22}H_{34}N_2NaO_6$ $(M+Na)^+$ 445.50, found 445.41

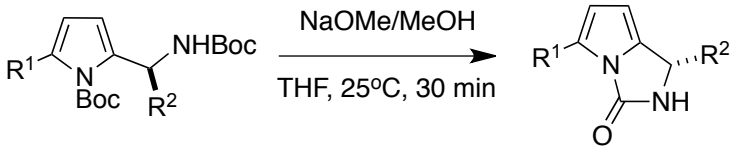


^{1}H NMR of **14f**

General Procedure and Preparation of Compounds 13a - e

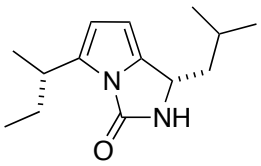
To a solution of the **12a - e** (1.0 equiv) in anhydrous THF (0.2 M) was added 6.5 M NaOMe/MeOH (5.0 equiv). The mixture solution was stirred for 30 min at 25 ^{o}C. The mixture was diluted with H_2O and then extracted with chloroform. The combined chloroform fractions were dried over $MgSO_4$ and concentrated under vacuum. The compound **13a - e** were purified by Flash chromatography.

Scheme S10. Synthesis of compounds **13a - e**.



12a - e **13a - e**

Preparation of Compound 13a



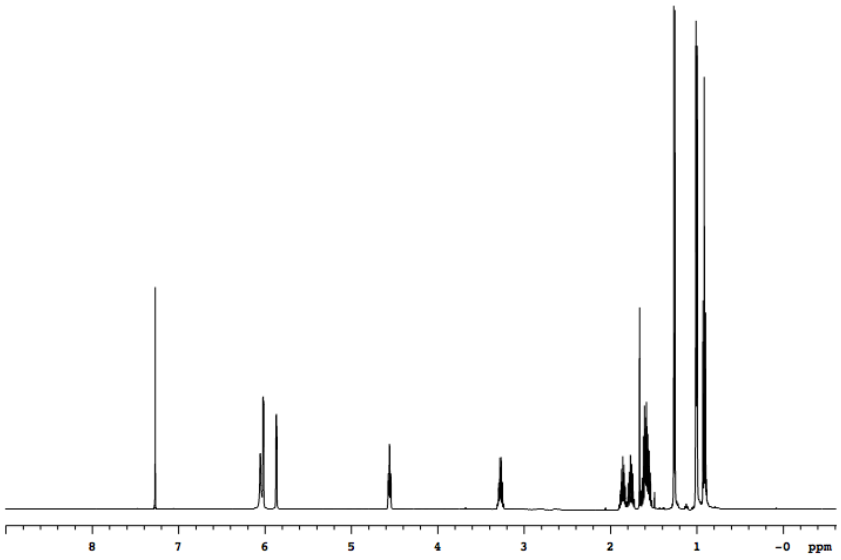
13a

Compound **13a** was prepared from **12a** (26.2 mg, 0.064 mmol). Flash chromatography (1:9 EtOAc/Hexanes) afforded 14.0 mg (93 %) **13a** as a yellowish oil.

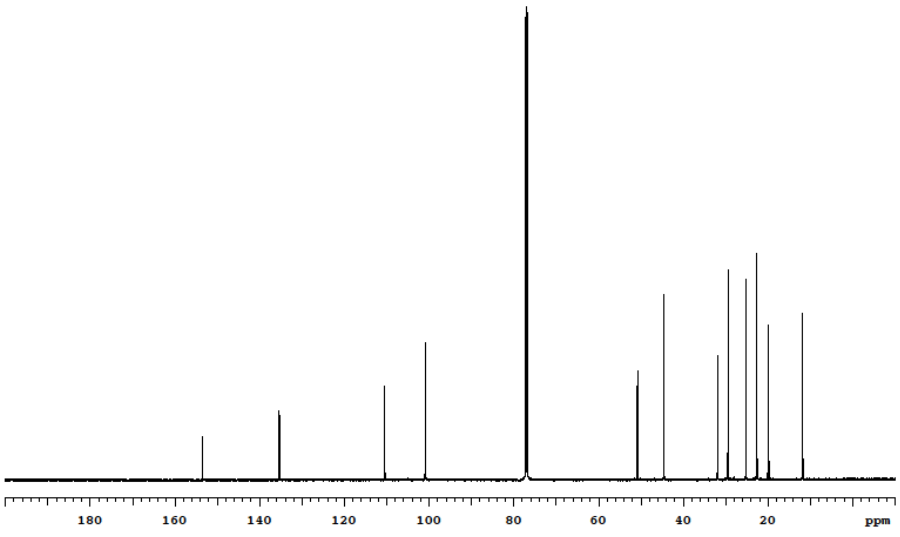
^{1}H NMR (500 MHz, $CDCl_3$) δ 6.05 (s, 1H), 6.01 (dd, 1H, J = 0.5, 1.5 Hz), 5.86 (dd, 1H, J = 1.5, 2.0 Hz), 4.55 (t, 1H, J = 6.5 Hz), 3.31-3.23 (m, 1H), 1.90-1.81 (m, 1H), 1.80-1.72 (m, 1H), 1.64- 1.51(m, 3H), 1.25 (d, 3H, J = 7.0 Hz), 1.00 (d, 3H, J = 2.5 Hz), 0.99 (d, 3H, J = 2.0 Hz), 0.91 (t, 3H, J = 7.5 Hz)

^{13}C NMR (125 MHz, $CDCl_3$) δ 154.4, 135.5, 135.2, 110.5, 100.9, 50.7, 44.5, 31.9, 29.4, 25.2, 22.6, 19.9, 11.7

MS (ESI, m/z) calcd for $C_{14}H_{23}N_2O$ $(M+H)^+$ 235.35, found 235.17

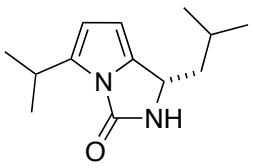


^{1}H NMR of **13a**



^{13}C NMR of **13a**

Preparation of Compound 13b



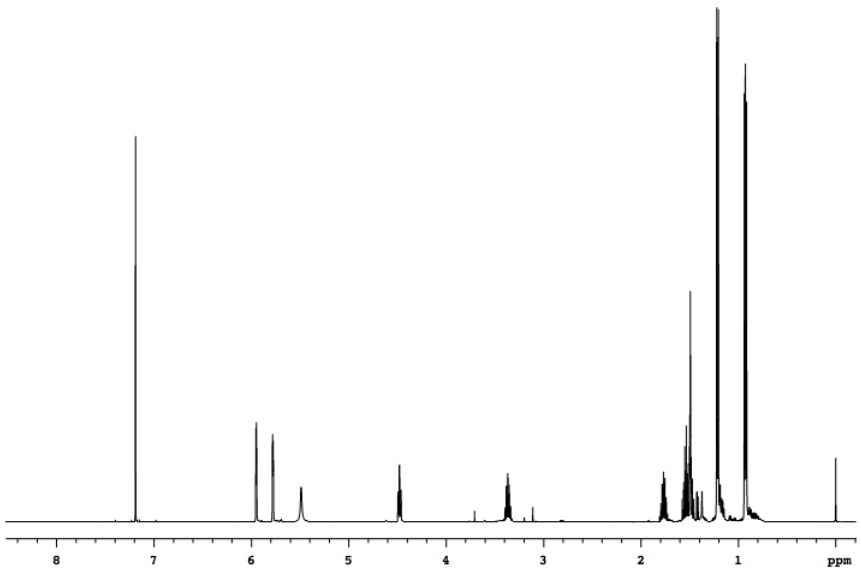
13b

Compound **13b** was prepared from **12b** (7.6 mg, 0.026 mmol). Flash chromatography (1:9 EtOAc/Hexanes) afforded 4.4 mg (99 %) **13b** as a yellowish oil.

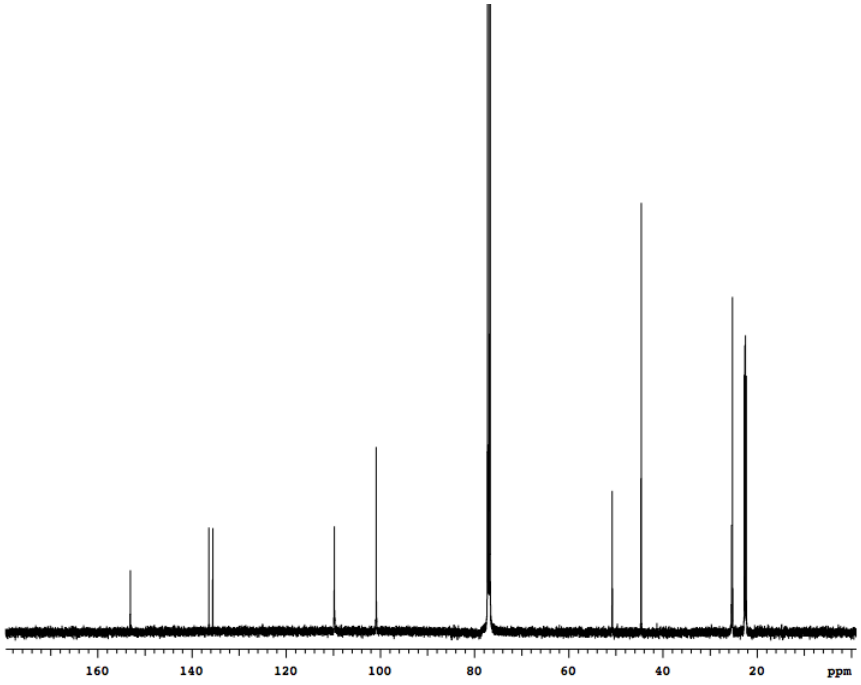
^{1}H NMR (500 MHz, $CDCl_3$) δ 6.02 (dd, 1H, J = 1.0, 2.5 Hz), 5.85 (dd, 1H, J = 2.5, 3.5 Hz), 5.56 (s, 1H), 4.55 (t, 1H, J = 7.0 Hz), 3.47-3.41 (m, 1H), 1.88-1.80 (m, 1H), 1.65-1.53 (m, 2H), 1.29 (d, 3H, J = 1.5 Hz), 1.28 (d, 3H, J = 1.5 Hz), 1.01 (d, 3H. J = 4.5Hz), 0.99 (d, 3H, J = 5.0 Hz)

^{13}C NMR (125 MHz, $CDCl_3$) δ 153.1, 136.4, 135.6, 109.8, 100.9, 50.8, 44.6, 25.4, 25.2, 22.7, 22.6, 22.5, 22.3

MS (ESI, m/z) calcd for $C_{13}H_{21}N_2O$ $(M+H)^+$ 221.16, found 221.17

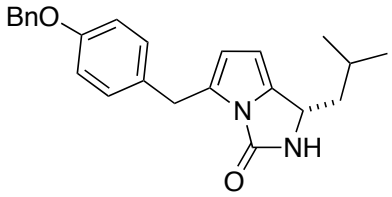


^{1}H NMR of **13b**



^{13}C NMR of **13b**

Preparation of Compound 13c



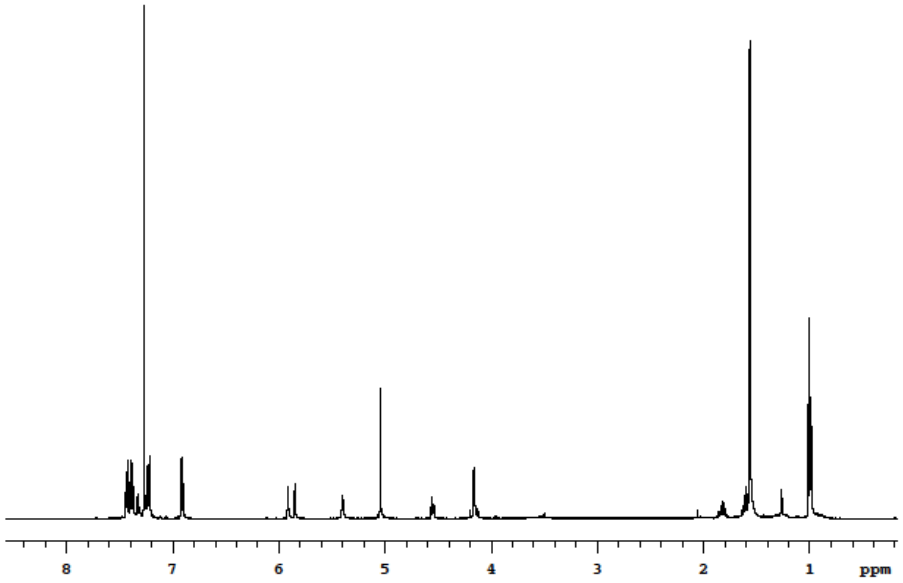
13c

Compound **13c** was prepared from **12c** (2.6 mg, 0.0047 mmol). Flash chromatography (1:7 EtOAc/Hexanes) afforded 1.8 mg (99 %) **13c** as an orange solid.

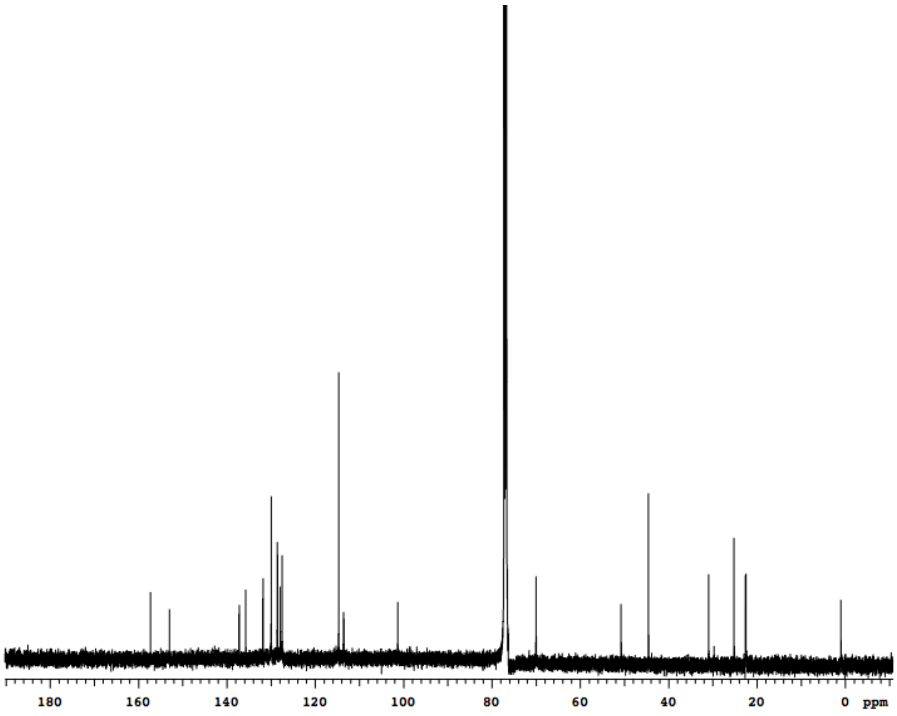
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.43 (d, 2H, J = 7.5 Hz), 7.38 (t, 2H, J = 7.0 Hz), 7.32 (t, 1H, J = 7.5 Hz), 7.22 (d, 2H, J = 8.5 Hz), 6.91 (d, 2H, J = 8.5 Hz), 5.91 (d, 1H, J = 3.0 Hz), 5.85 (dd, 1H, J = 1.0, 8.0 Hz), 5.40 (s, 1H), 5.04 (s, 2H), 4.55 (t, 1H, J = 7.0 Hz), 4.16 (d, 2H, J = 5.0 Hz), 1.85-1.80 (m, 1H), 1.64-1.52 (m, 2H), 0.99 (t, 6H, J = 6.5 Hz)

^{13}C NMR (125 MHz, $CDCl_3$) δ 157.3, 153.0, 137.2, 135.7, 131.8, 129.9, 128.6, 128.5, 127.9, 127.5, 114.7, 113.5, 101.3, 70.0, 50.8, 44.6, 31.0, 25.2, 22.6, 22.5

MS (ESI, m/z) calcd for $C_{24}H_{27}N_2O_2$ $(M+H)^+$ 375.21, found 375.16

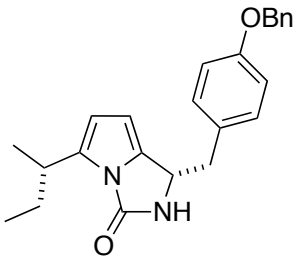


^{1}H NMR of **13c**



^{13}C NMR of **13c**

Preparation of Compound 13d



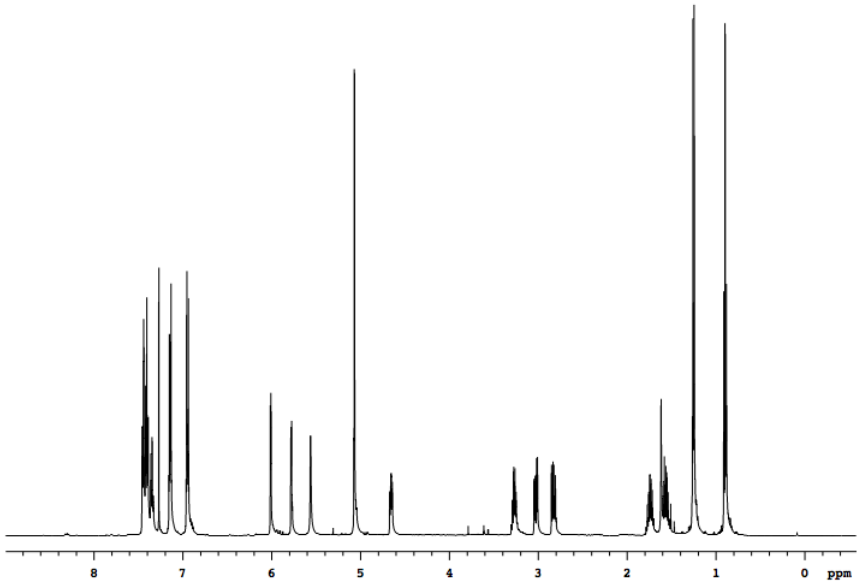
13d

Compound **13d** was prepared from **12d** (31.1 mg, 0.054 mmol). Flash chromatography (1:7 EtOAc/Hexanes) afforded 13.1 mg (77 %) **13d** as a yellowish oil.

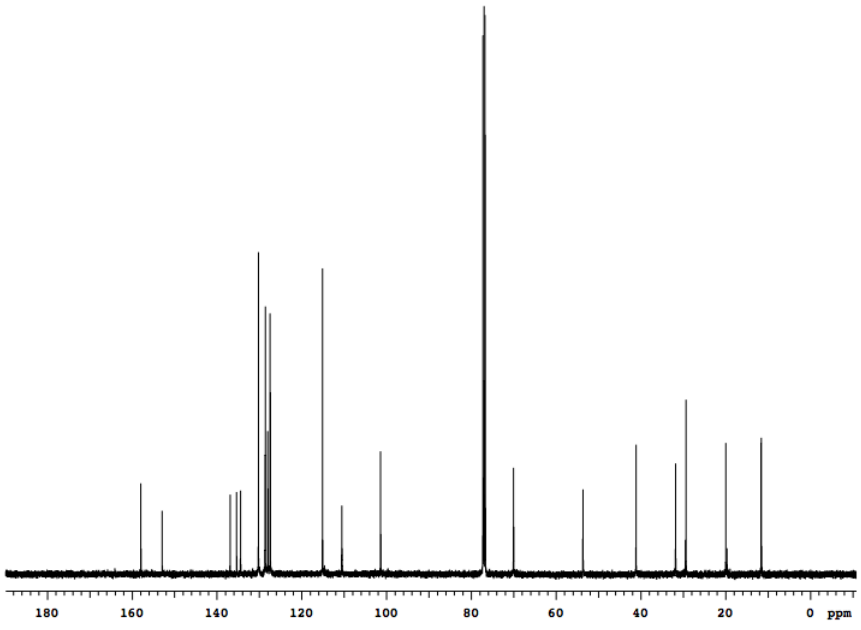
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.45 (d, 2H, J= 7.0 Hz), 7.41 (t, 2H, J = 7.5 Hz), 7.35 (t, 1H, J = 7.5 Hz), 7.14 (d, 2H, J = 8.5 Hz), 6.94 (d, 2H, J = 9.0 Hz), 6.01 (d, 1H, J = 3.0 Hz), 5.78 (dd, 1H, J = 1.5, 3.0 Hz), 5.56 (s, 1H), 4.66 (t, 1H, J = 7.5 Hz), 3.30-3.24 (m, 1H), 3.03 (dd, 1H, J = 5.5, 13.5 Hz), 2.83 (dd, 1H, J = 8.5, 14.0 Hz), 1.79-1.70 (m, 1H), 1.62-1.52 (m, 1H), 1.25 (d, 3H, J = 7.0 Hz), 0.90 (t, 3H, 7.0 Hz)

^{13}C NMR (125 MHz, $CDCl_3$) δ 157.9, 152.9, 136.9, 135.4, 134.4, 130.2, 128.7, 128.6, 128.0, 127.5, 115.1, 110.5, 101.4, 70.0, 53.7, 41.2, 31.8, 29.4, 20.0, 11.6

MS (ESI, m/z) calcd for $C_{24}H_{27}N_2O_2$ $(M+H)^+$ 375.21, found 375.21

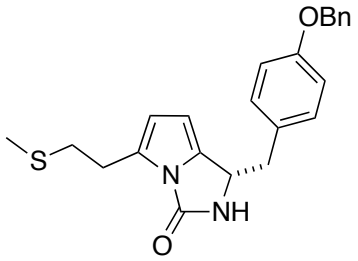


^{1}H NMR of **13d**



^{13}C NMR of **13d**

Preparation of Compound 13e



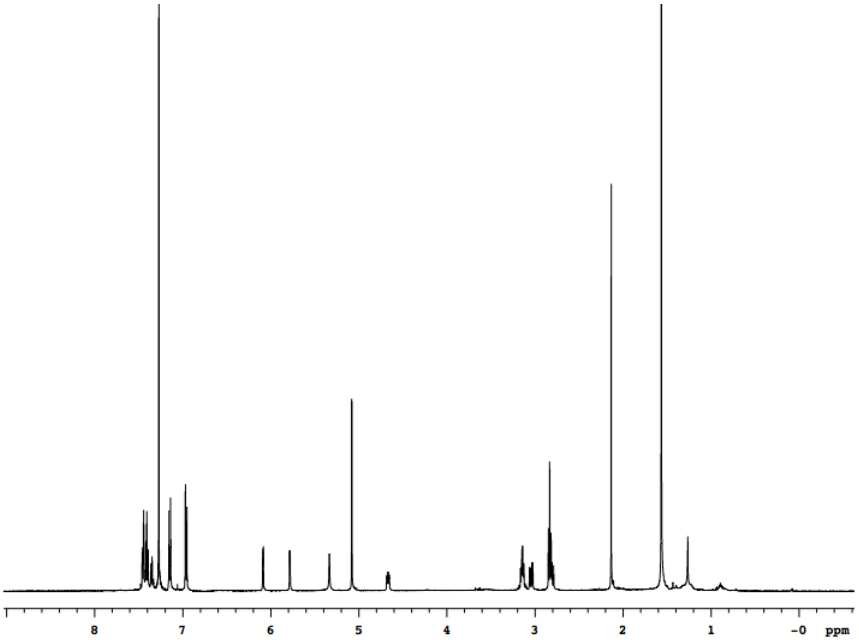
13e

Compound **13e** was prepared from **12e** (2.7 mg, 0.0048 mmol). Flash chromatography (1:9 EtOAc/Hexanes) afforded 1.5 mg (94 %) **13e** as a white solid.

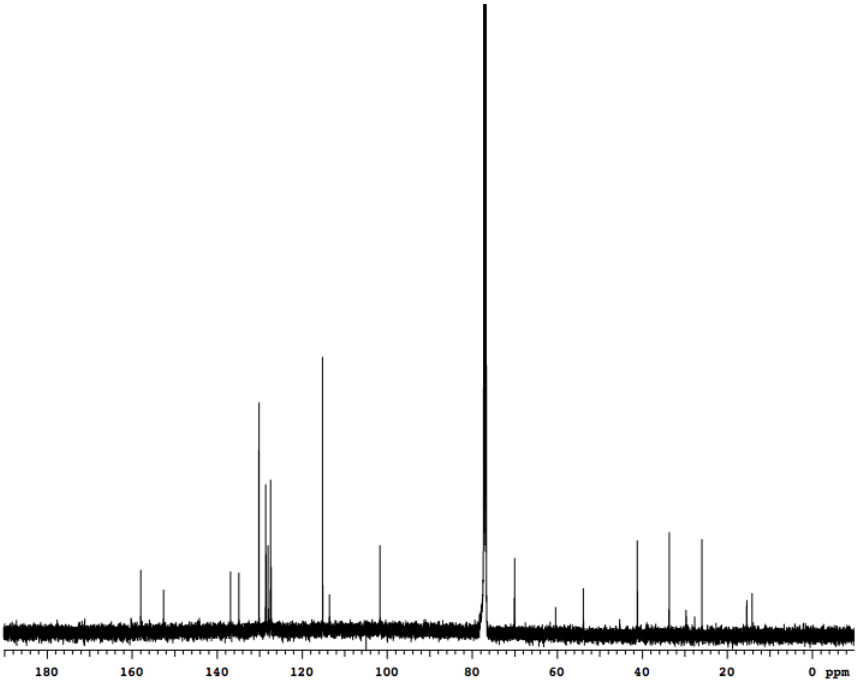
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.44 (d, 2H, 7.0 Hz), 7.40 (t, 2H, J = 7.0 Hz), 7.34 (t, 1H, J = 7.5 Hz), 7.14 (d, 2H, J = 9.0 Hz), 6.95 (d, 2H, J = 8.5 Hz), 6.08 (d, 1H, J = 3.5 Hz), 5.77 (dd, 1H, J = 1.5, 3.0 Hz), 5.33 (s, 1H), 5.07 (s, 2H), 4.66 (t, 1H, 8.0 Hz), 3.17-3.12 (m, 2H), 3.03 (dd, 1H, J = 5.5, 14.0 Hz), 2.83 (t, 2H, 7.5 Hz), 2.80 (dd, 1H, J = 9.0, 15.0 Hz), 2.13 (s, 3H)

^{13}C NMR (125 MHz, $CDCl_3$) δ 158.0, 152.6, 136.8, 134.9, 130.2, 128.6, 128.6, 128.0, 127.5, 127.3, 115.2, 113.6, 101.7, 70.1, 53.9, 41.2, 33.7, 26.0, 15.4

MS (ESI, m/z) calcd for $C_{23}H_{25}N_2O_2S$ $(M+H)^+$ 393.16, found 393.13



^{1}H NMR of **13e**

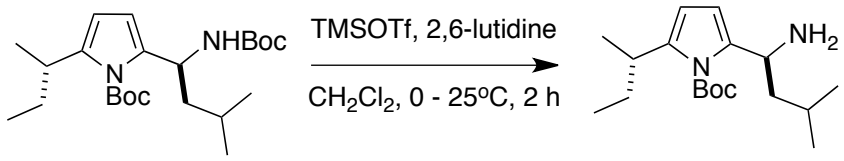


^{13}C NMR of **13e**

General Procedure and Preparation of Compound 3a

To a solution of the **12a** (1.0 equiv) and 2,6-lutidine (5.0 equiv) in anhydrous CH_2Cl_2 (0.1 M) and was added TMSOTf (4.0 equiv) at 0 ^oC. The mixture solution was stirred for 15 min at 0 ^oC. The mixture was warmed up to 25 ^oC and then stirred for 2 h at 25 ^oC. To the mixture was added Sat. NH_4Cl (aq) and then extracted with EtOAc. The combined organic layer was dried over $MgSO_4$ and concentrated under vacuum. The compound was purified by Flash chromatography.

Scheme S11. Synthesis of compounds **3a**.



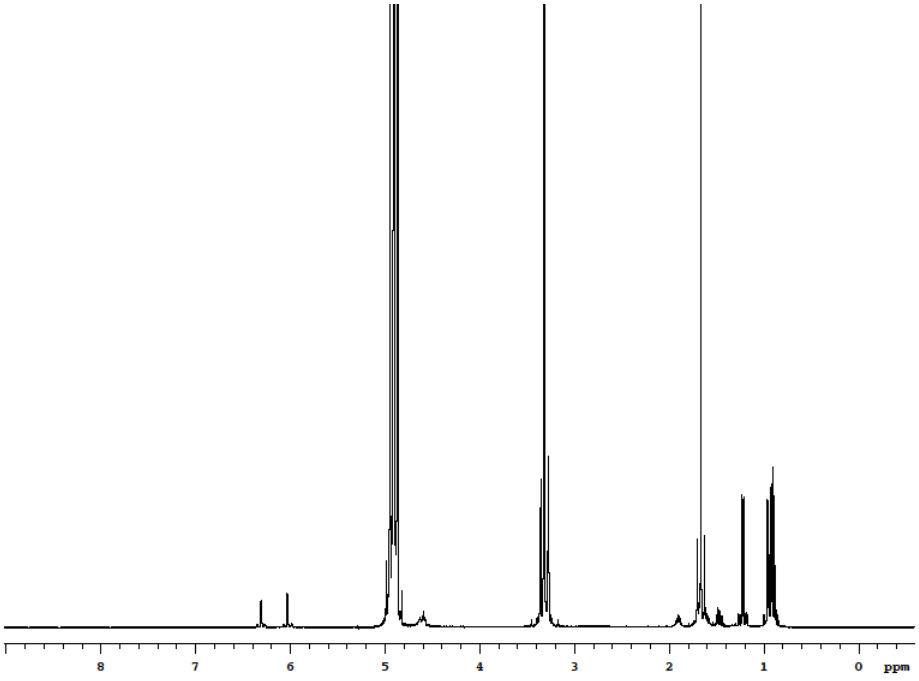
12a **3a**

Compound **3a** was prepared from **12a** (3.0 mg, 0.0073 mmol). Flash chromatography (1:1 EtOAc/Hexanes) afforded 1.5 mg (67 %) **3a** as a colorless oil.

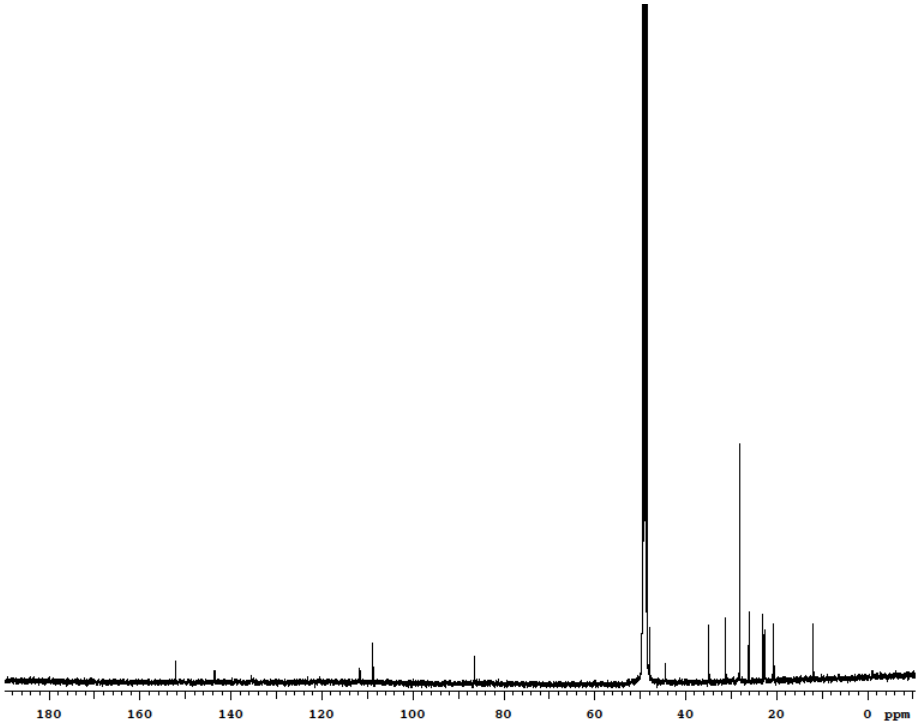
^1H NMR (500 MHz, CD_3OD) δ 6.31 (d, 1H, J = 3.5 Hz), 6.03 (d, 1H, J = 3.5 Hz), 4.70-4.62 (m, 1H), 4.59 (t, 1H, J = 8.5 Hz), 1.92-1.88 (m, 1H), 1.74-1.70 (m, 1H), 1.66 (s, 9H), 1.64-1.58 (m, 2H), 1.50-1.44 (m, 1H), 1.22 (d, 3H, J = 6.5 Hz), 0.95 (d, 3H, J = 7.0 Hz), 0.92 (d, 3H, J = 7.0 Hz), 0.89 (d, 3H, J = 7.5 Hz)

^{13}C NMR (125 MHz, CD_3OD) δ 152.2, 143.6, 135.6, 111.6, 108.8, 86.4, 48.0, 44.5, 34.9, 31.3, 28.2, 26.1, 23.0, 22.7, 20.6, 11.9

MS (ESI, m/z) calcd for $C_{18}H_{33}N_2O_2$ $(M+H)^+$ 309.25, found 309.26



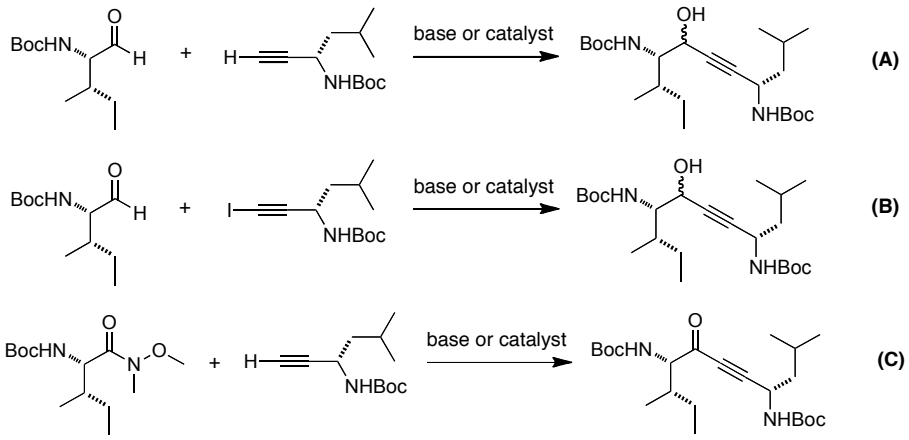
^{1}H NMR of **3a**

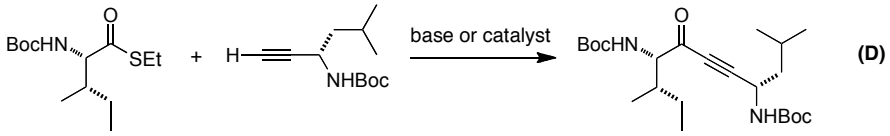


^{13}C NMR of **3a**

Optimization of Reactions

Table S1. Screening of Reaction Conditions for the Formation of Propargyl Alcohols or Propargyl Ketones.

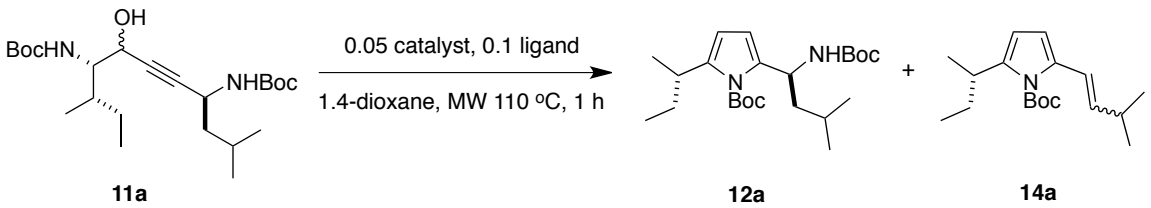




entry	reaction[a]	base or catalyst	solvent	temp. (°C)	yield (%)[j]
1[b]	A	LHMDS	THF	-78	57
2[c]	A	^{n}BuLi	THF	-78	NR
3[d]	A	^{n}BuLi + HMPA	THF	-78	36
4[e]	A	^{n}BuLi + diisopropylamine	THF	-78	40
5[f]	A	EtMgBr	THF	0	34
6[g]	B	$CrCl_2$+$NiCl_2$	THF/DMF (1:1)	25	NR
7[h]	C	^{n}BuLi	THF	-78	NR
8[i]	D	$PdCl_2$(dppf) + P(2-furyl)$_3$ + CuI	DMF/TEA (5:1)	50	21

[a] Aldehyde (1.0 equiv), Weinreb amide (1.0 equiv), thioester (1.0 equiv), terminal alkyne (2.5 equiv), and iodoalkyne (2.0 equiv) were used. [b] LHMDS (6.5 equiv). [c] ^{n}BuLi (6.5 equiv). [d] ^{n}BuLi (6.3 equiv) and HMPA (12.0 equiv). [e] ^{n}BuLi (6.5 equiv) and diisopropylamine (6.5 equiv). [f] EtMgBr (6.5 equiv). [g] $CrCl_2$ (10.0 equiv) and $NiCl_2$ (0.1 equiv). [h] ^{n}BuLi (6.5 equiv). [i] $PdCl_2$(dppf) (0.1 equiv), P(2-furyl)$_3$ (0.25 equiv) and CuI (1.7 equiv). [j] isolated yields.

Table S2. Screening of Reaction Conditions for Hydroamination.



11a **12a** **14a**

entry	catalyst	ligand	conversion yield (%)[a]	**12a**:**14a**[a]
1	$PtCl_2$	-	94	6.8:1.0
2	$Pd(OAc)_2$	-	13	1.0:trace
3	$PdCl_2(PPh_3)_2$	-	10	1.0:trace
4	$PtCl_2$	citric acid	87	5.8:1.0
5	$PtCl_2$	acetic acid	90	5.9:1.0
6	$PtCl_2$	K_2CO_3	87	5.6:1.0
7	$PtCl_2$	Cs_2CO_3	70	6.0:1.0
8	$PtCl_2$	PPh_3	90	5.3:1.0
9	$PtCl_2$	(2-biphenyl)di-tert-butylphosphine	88	6.0:1.0

[a] determined by NMR.

APPENDIX F

EXPERIMENTAL FOR CHAPTER VI

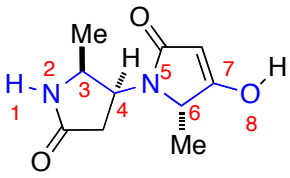
A. Templates For Secondary Structures

Standard template for overlays with 3_{10}-helix, α-helix, π-helix, and β-strand were obtained from Discovery Studio 2.5. Parallel β-sheet and sheet/turn/sheet templates were obtained by modified β-sheet builder.
(http://www-lbit.iro.umontreal.ca/bBuilder/index.html)

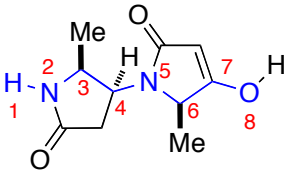
B. Procedure For Overlays

After minimization in the QMD process, the conformers were grouped into families based on their Cα-Cβ coordinates. The process of systematically matching preferred conformers with secondary structures was performed in the following way. All the conformers within 3.0 kcal/mol were considered to be "preferred". Each of these was overlaid on ideal secondary structures using an in house generated algorithm that compared Cα-Cβ coordinates of the side chains which generates a list of structures ranked in terms of the RMSD for the overlay process. The lowest energy structures from each family were each overlaid on the ideal secondary structures shown in chapter 6, Table 6.1 and Table 6.2.

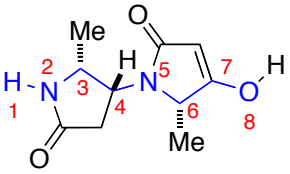
C. Procedure For Ramachandran Plot



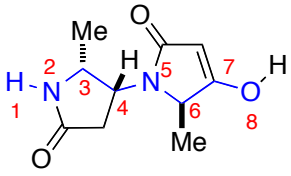
LL



LD



DL

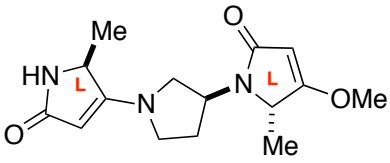


DD

Definition of torsional angles: ϕ (H^1-N^2-C^3-C^4), ψ (N^2-C^3-C^4-N^5), ϕ' (C^4-N^5-C^6-C^7) and ψ' (N^5-C^6-C^7-O^8)

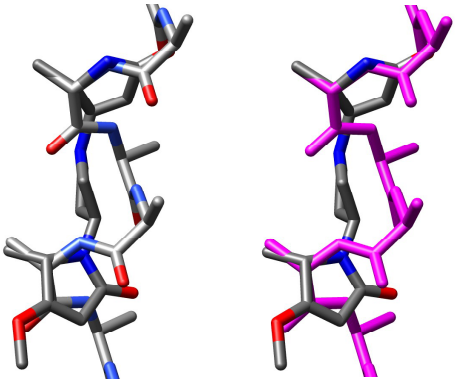
For each compound, 600 conformers within 3.0 kcal/mol were obtained by QMD. Ramachandran plot for the conformers was obtained from UCSF Chimera 1.5.2. www.cgl.ucsf.edu/chimera/

D. Overlays of Compound 6 Conformers On The Secondary Structures

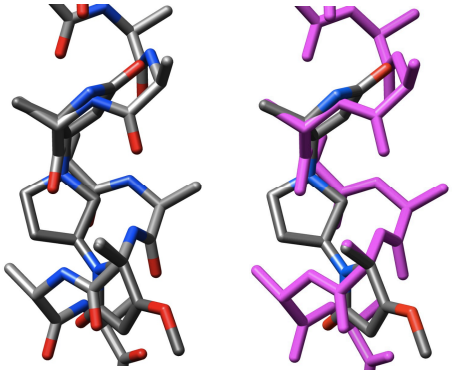


6

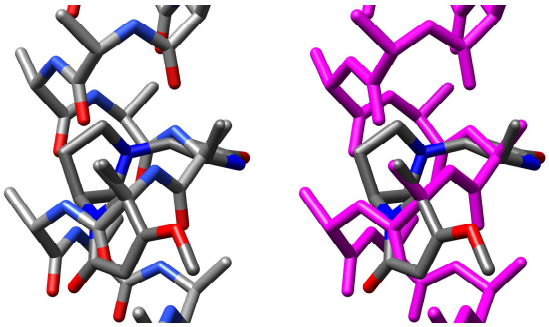
overlay on 3_{10}-helix



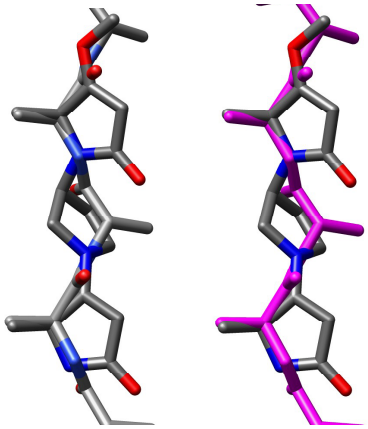
overlay on α-helix



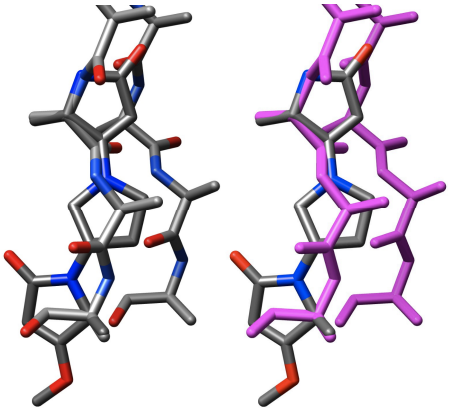
overlay on π-helix



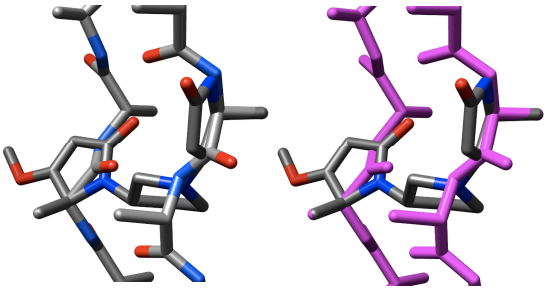
overlay on β-strand



overlay on parallel β-sheet



overlay on sheet/β-turn/sheet



APPENDIX G

EXPERIMENTAL FOR CHAPTER VII

A. Enterotoxin and Cholera Toxin as Targets for Pyrrolinone-pyrrolidine Oligomers 1

Structure and Function of Cholera Toxin and Enterotoxin

These toxins are directly associated with causing cholera and related enteropathies in humans and domestic animals. Diarrhea is perhaps the leading worldwide cause of mortality for children under five, and the featured toxins are responsible for a significant fraction of these.[164]

Structurally, both toxins consist of a 27 kDa A fragment which sits on top of a cyclic homopentamer of 11.7 kDa B fragments giving an AB_5 quaternary arrangement.[165] Over 80 % of both the A and B fragments in the two toxins share the same amino acid sequence.[165] Figure S1 shows side and front views of *E. coli* enterotoxin from a structure pdb:1eef;[166] template **1** (LLL) overlaid with a region of this structure is just visible in these pictures, and expansions are given below.

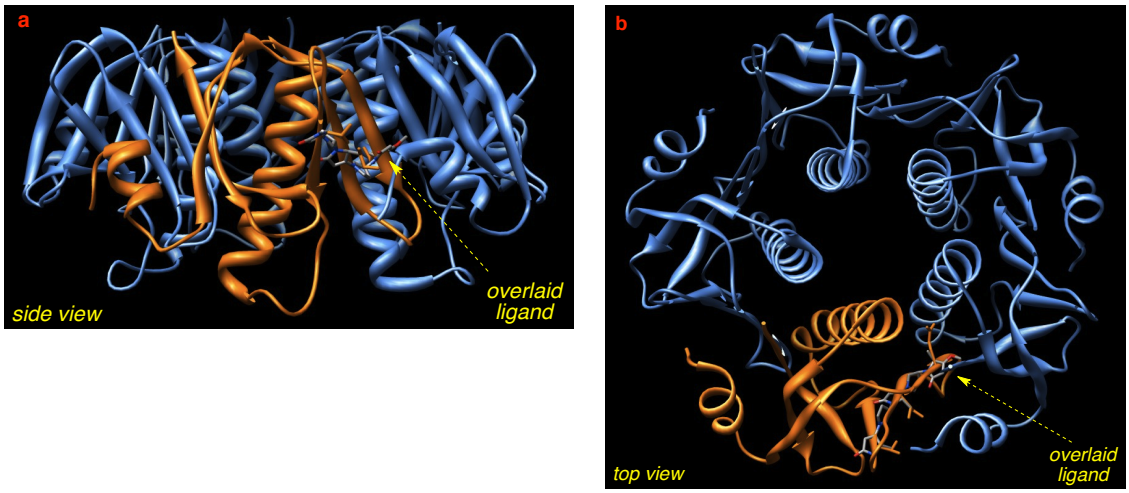


Figure S1. Side and front views of the E. coli enterotoxin from PDB:1eef.

In the progression of the parent diseases, bacterial cells express the constituent A and B fragments, and these assemble into the AB_5 hexamer units. It is the B_5 units of the AB_5 structures that bind the ganglioside GM1 receptor of the host's epithelial cells. Binding of the B_5 pentamer unit triggers down-regulation of pro-inflammatory immune responses.[164,167] Receptor-mediated endocytosis delivers the toxin into the cells, and then the A unit is proteolytically cleaved. This fragment catalyzes ADP ribosylation of the $G\alpha_s$ subunit of the heterotrimeric G protein resulting in constitutive cAMP production, secretion of water and salts into the lumen of the small intestine resulting in rapid dehydration and other factors associated with cholera.

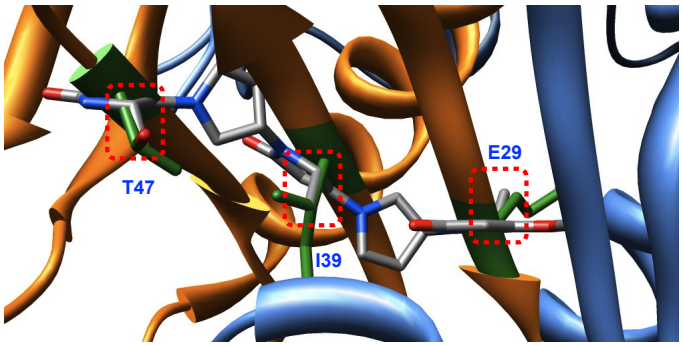
Each B-unit in the B_5 structure shares an extended protein-protein interface, making the pentamers extremely stable. They maintain their secondary structures in ionic detergents, 8.0 M urea, 7.0 M guanadinium hydrochloride, and to temperatures in excess of 80 °C in aqueous solution. Correspondingly, high activation energy (151 kJ/mol) has been measured for disassembly on the pentamer units, but they can be denatured into monomeric fragments at pH 2 or less. Dissociated toxins fragments at low pH assemble at experimentally convenient rates once the medium is made neutral again.

We propose that a small molecule interface mimic is unlikely to be able to cleave the preformed AB_5 hexamers of these toxins, since they are so stable. However, the impact of the mimics could be assayed in vitro by monitoring their effect on their rate of re-assembly after pH reduction then restoration to neutrality. A possible therapeutic mode of action for compounds that suppress assembly of the B_5 units would be via penetration of the small molecule into bacterial cells in the gut of the host, preventing expression and formation of the mature hexamer before it is released.

Data From EKO and eKO

Mining of the database using EKO for stereomer **1** (LLL) gave thirteen different results for cholera/enterotoxins within RMSD 0.33 Å (1eef, 1b44, 1lts, 1lt4, 1fd7, 1djr, 1s5c, 1chq, 1pzi, 1lt6, 1ltg, 1ltt, 1lti). In all results, conformers of stereomer **1** were on Thr47-Ile39-Glu29 residues set of B_5 units. Figure S1 shows an overlay for the best matching from EKO. The conformer that is 2.89 kcal/mol over the lowest energy conformation was matched with RMSD 0.26 on Thr47-Ile39-Glu29. Application of eKO to the enterotoxin crystal structure 1eef gave Met31-Glu29-Tyr27 (*RMSD* 0.42 Å, ΔG^o = 1.78 kcal/mol, Figure S2 a) and Lys102-Ser100-Ala98 (*RMSD* 0.49 Å, ΔG^o = 1.78 kcal/mol, Figure S2 b). The Met31-Glu29-Tyr27 residues set correspond to a region of the interface that is known to be vital for H-bonding.[7]

a



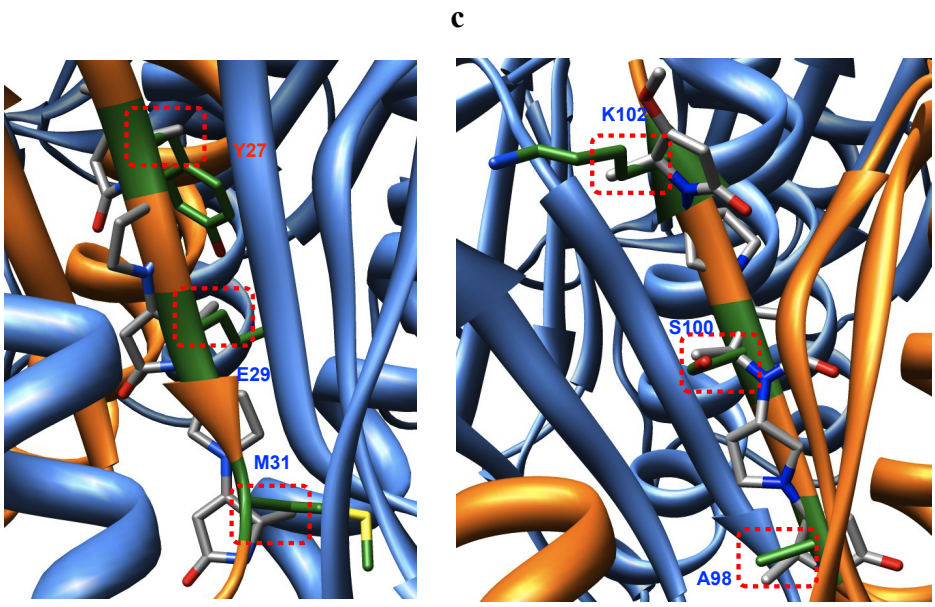


Figure S2. (a) Mining 1 (L,L,L) gave a good overlay on the T47-I39-E29 from EKO; (b) relaxing the RMSD requirement exposed two other matches (b) and (c) at Met31-Glu29-Tyr27 and Ala98-Ser100-Lys102 from eKO.

Structures of the featured toxins were mined for all 8 amino acid-based epimers of **1**. Twenty hits were observed for **1** (LLL), and all of them matched on the same protein region T47-I39-E29. Only **1** (LDL) of the other isomers matched, and this time with I31-E29-Y27 of the cholera toxin, which corresponds to the "second tier" hit (M31-E29-Y27) for matching **1** (LLL) on the enterotoxin (5 hits); all the matches were *C*-to-*N*.

Table S1. Matching of Stereoisomers of 1 on Toxin Structures.

	conformer	PDB	RMSD (Å)	score	residues	direction	source	# of conformers
1	LLL	1eef	0.26	13.2	T47-I39-E29	*C* -> *N*	E.Coli	2
2	LLL	1b44	0.26	13.5	T47-I39-E29	*C* -> *N*	E.Coli	2
3	LLL	1lts	0.27	14.3	T47-I39-E29	*C* -> *N*	E.Coli	2
4	LLL	1djr	0.28	15.0	T47-I39-E29	*C* -> *N*	E.Coli	2
5	LLL	1fd7	0.30	14.0	T47-I39-E29	*C* -> *N*	E.Coli	1
6	LLL	1efi	0.30	15.2	T47-I39-E29	*C* -> *N*	E.Coli	2
7	LLL	1pzi	0.30	15.7	T47-I39-E29	*C* -> *N*	E.Coli	4
8	LLL	1g8z	0.30	16.0	T47-I39-E29	*C* -> *N*	E.Coli	1
9	LLL	1s5c	0.31	13.9	T47-I39-E29	*C* -> *N*	cholera	1
10	LLL	1chq	0.31	13.9	T47-I39-E29	*C* -> *N*	cholera	1
11	LLL	1lt4	0.31	14.1	T47-I39-E29	*C* -> *N*	E.Coli	2
12	LLL	1ct1	0.31	15.0	T47-I39-E29	*C* -> *N*	cholera	1

13	LLL	1lt5	0.31	15.2	T47-I39-E29	*C* -> *N*	E.Coli	1
14	LLL	1efi	0.31	15.5	T47-I39-E29	*C* -> *N*	E.Coli	1
15	LLL	1lt6	0.31	15.9	T47-I39-E29	*C* -> *N*	E.Coli	2
16	LLL	1s5f	0.31	16.1	T47-I39-E29	*C* -> *N*	cholera	1
17	LLL	1lta	0.31	16.2	T47-I39-E29	*C* -> *N*	E.Coli	1
18	LLL	1s5e	0.31	16.2	T47-I39-E29	*C* -> *N*	cholera	1
19	LLL	1ltg	0.32	16.1	T47-I39-E29	*C* -> *N*	E.Coli	2
20	LLL	1lti	0.33	16.4	T47-I39-E29	*C* -> *N*	E.Coli	1
21	LDL	1g8z	0.27	18.1	L31-E29-Y27	*C* -> *N*	cholera	1
22	LDL	1md2	0.29	18.3	L31-E29-Y27	*C* -> *N*	cholera	3
23	LDL	1s5c	0.29	20.3	L31-E29-Y27	*C* -> *N*	cholera	1
24	LDL	1pzj	0.30	18.2	L31-E29-Y27	*C* -> *N*	cholera	1
25	LDL	1rdp	0.30	18.6	L31-E29-Y27	*C* -> *N*	cholera	1

B. AICAR as Targets for Pyrrolinone-pyrrolidine Oligomers 1

Structure and Function of AICAR

5-Aminoimidazole-4-carboxamide ribonucleotide transformylase (AICAR Tfase) is one component of a bifunctional enzyme, the other being inosine 5'-monophosphate cyclohydrolase (IMPCH).[169,170] These catalyze the last two steps in purine biosynthesis (Figure S3).[171,172] Formyl transfer from the cofactor 10-formyl-tetrahydrofolate (10-f-THF) to the aminoimidazole functionality is mediated by AICAR Tfase, then IMPCH promotes cyclization of this *N*-formyl group to give the purine framework (of IMP). Normal cells generate most of the purine they require by a salvage pathway; for them, *de novo* biosynthesis as illustrated in Figure S3 is relatively unimportant.[173-176] However, cancer cells depend heavily on the *de novo* pathway, hence they are vulnerable to inhibitors.

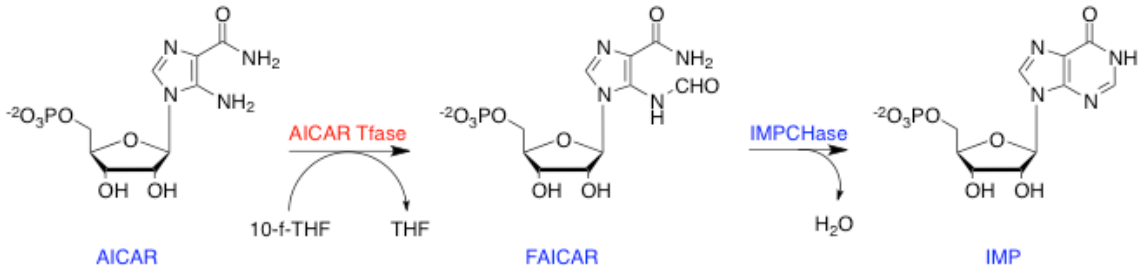


Figure S3. The bifunctional enzyme ATIC completes purine syntheses: formyl transfer mediated by AICAR Tfase, then ring closure by IMPCHase.

AICAR Tfase is one of several folate-dependent enzyme targets for chemotherapy {*cf* thymidylidyl synthase, dihydrofolate reductase, and glycinamide ribonucleotide transformylase}. Inhibitors of AICAR Tfase that are *not* based on folate have advantages over other *anti*-folate drugs (*cf* DHFR, "DDATHF" {lometrexol}[177,178] and LY231514[179]) because they are unlikely to impact non-targeted folate-dependent enzymes giving unpredictable side effects. Two validated strategies for disabling AICAR Tfase that do *not* involve mimicry of folate are: (i) disruption of the active site function; and, (ii) perturbation the interface in the dimer. AICAR Tfase is only active in the dimeric form, and some molecules that disrupt the dimer interface are known to inhibit the enzyme. These molecules are cyclic peptides (K_i 17 μM or more),[180] or flexible small molecules (from HTS, e.g. K_i 17 μM).[181] Thus the only approaches used so far to disrupt AICAR Tfase dimerization have been combinatorial, involving large numbers of randomly produced compounds; they give relatively weak inhibition.

Data From EKO and eKO

The best hit for screening **1** (LLL) had the sequence Ala218-Ley220-Thr222; these align with a sheet region on the interface (RMSD 0.28 and 0.31, Figure S5a). Relaxing the RMSD requirement gave three more hits (RMSD 0.39, 0.40, and 0.42, Figure S5c - d). Mining all the amino acid-based stereoisomers of **1** gave only one more hit, for **1** (LDL) and this matched a different region of the protein Lue329-Glu331-Lys333. All the mimics aligned parallel with the strand.

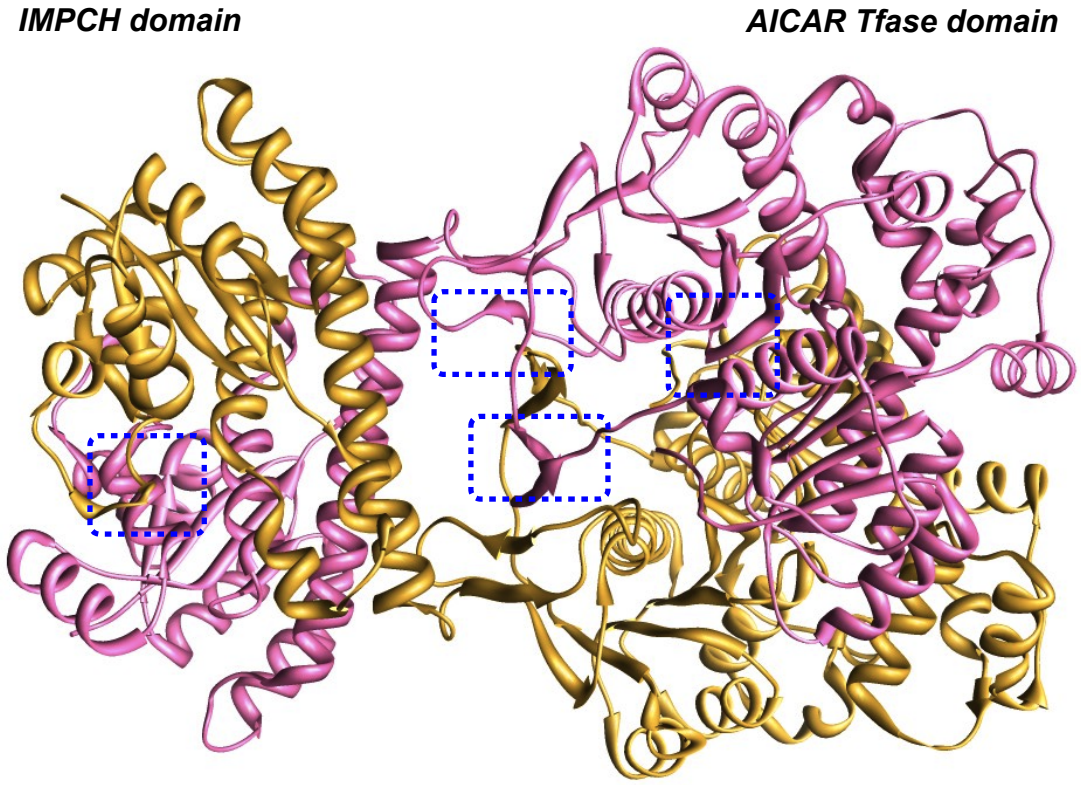
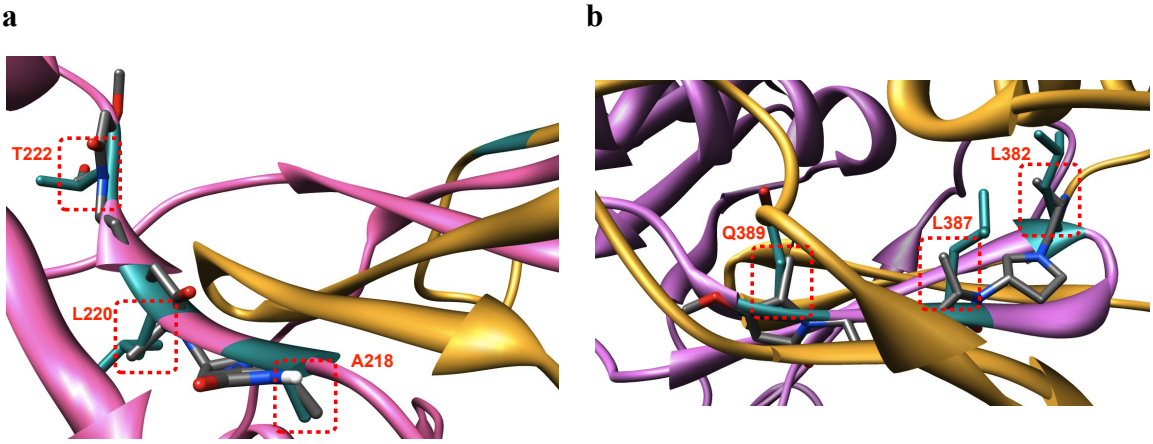


Figure S4. Structure of the bifunctional enzyme with key interface regions where scaffold 1 overlays well highlighted.



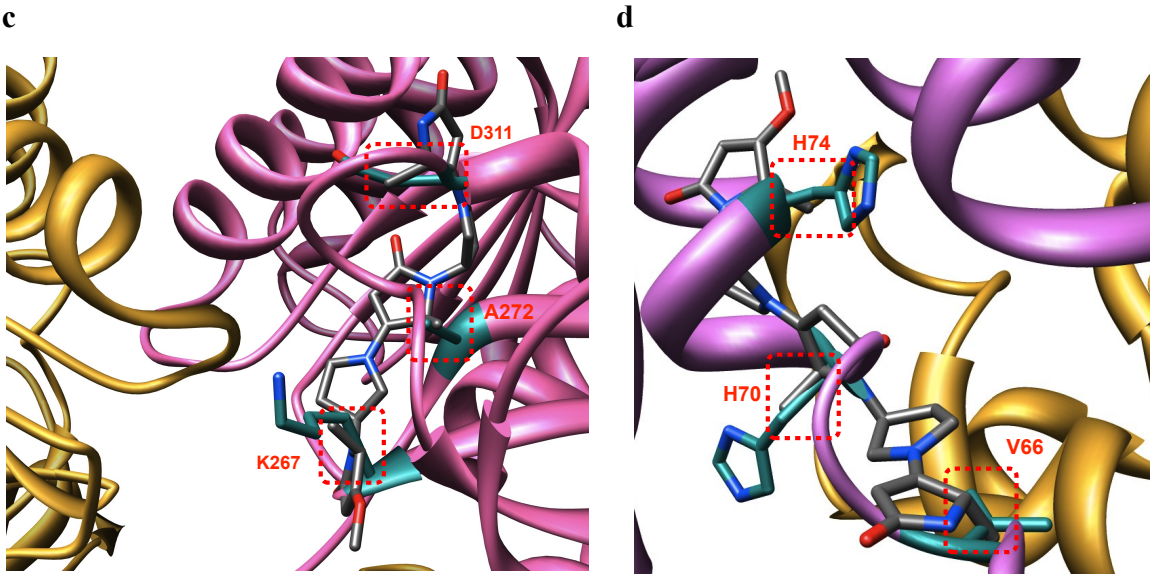


Figure S5. (a) – (d) Expansions of these regions with key side-chain residues shown in the red boxes, and with preferred conformations of scaffold **1** overlaid. Key comparisons are for the Cα and Cβ atoms. There have been no thermodynamic analyses of AICAR mutants to determine hot-spots.

Table S2. Matching of Stereoisomers of **1** on AICAR Structures.

	conformer	PDB	RMSD (Å)	score	residues	direction	source
1	LLL	1thz	0.28	17.6	A218-L220-T222	*N* -> *C*	chicken
2	LLL	1m9n	0.31	19.3	A218-L220-T222	*N* -> *C*	chicken
3	LDL	1zcz	0.26	17.0	L329-E331-K333	*N* -> *C*	thermotoga maritima

C. GAPDH as Targets for Pyrrolinone-pyrrolidine Oligomers 1

Structure and Function of GAPDH

D-GlycerAldehyde-3-Phosphate DeHydrogenase (GAPDH) mediates oxidative phosphorylation of the aldehyde after which it is named; this is a key step in the glycolytic pathway (Figure S6a).[182] The structure of GAPDH is a homotetramer or, more accurately, a dimer of dimers, wherein the active site is a NAD^+ binding groove found on each monomer component (Figure S6b and c).

a

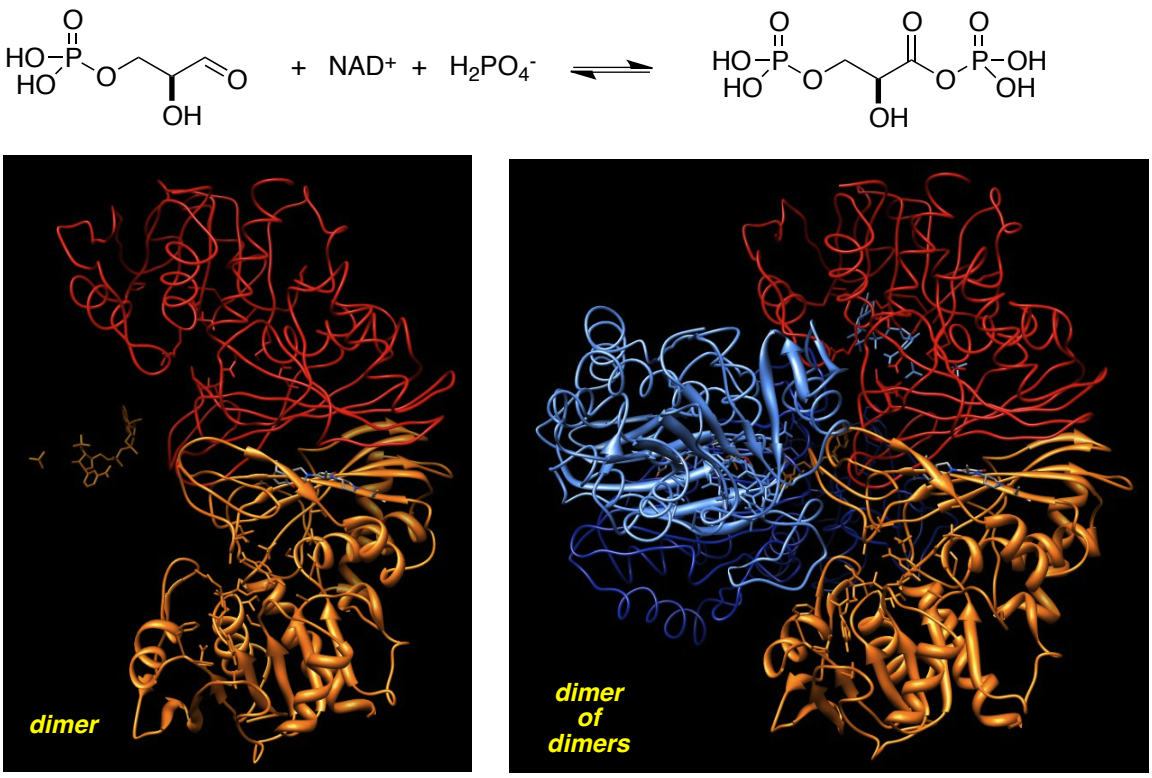


Figure S6. (a) Normal glycolytic function of GAPDH; (b) dimer of GAPDH; and (c) dimers of GAPDH associate to give a tetrameric quaternary structure.

More sinister roles, however, have been attributed to GAPDH, because this enzyme is implicated in apoptotic cell death, particularly in neurodegeneration. Thus, in cellular assays, rescue from apoptosis can be affected by antisense suppression of GAPDH or using the Parkinson's therapy (*R*)-deprenyl (Selegiline).[183,184] Further, a tricyclic deprenyl analog, CGP3466, binds and stabilizes the dimeric form of GAPDH and has 100x the rescuing effect of deprenyl *in vitro*; CGP3466 is a neuroprotective drug that has featured in clinical trials for Parkinson's disease and ALS. Consistent with these observations, certain fractions of cerebrospinal fluid (CSF) from Parkinson's patients cause apoptosis when added to cells in culture, whereas CSF from healthy patients does not. Further, the apoptidic effects of CSF from Parkinson's patients is prevented by antisense targeting of GAPDH or by (*R*)-deprenyl.[185,186]

Based on the assertions above, nefarious roles of GAPDH in several neurodegenerative diseases are implicated. *Exact* mechanisms that tie GAPDH to apoptosis in neurological diseases like Huntington's, Parkinson's, Alzheimer's, ALS, stroke and glaucoma (neurodegeneration of retinal ganglion cells) are not known, but this has been an area of intense recent interest (GAPDH in human neurodegenerative diseases has been reviewed)[185-187] and some clues are emerging.

GAPDH has to be imported into the nucleus to trigger apoptosis. After this, nuclear accumulation of GAPDH, or an isoform of it,[188] occurs in the neurological diseases mentioned so far. Association of cystolic GAPDH with the E3 ubiquitin ligase Siah 1 is critical for importing the former into the nucleus, because only the latter has a nuclear localization signal.[189] It has been proposed that CGP3466 may bind the NAD^+ site causing structural changes that reduce the affinity of GAPDH for Siah 1; in other words, the drug inhibits apoptotic activity of GAPDH by preventing its nuclear localization.[190] Precisely what form of nuclear GAPDH triggers apoptosis is unclear; some evidence suggests that GAPDH-complex stabilizes the otherwise short-lived Siah 1,[189] but another explanation is that activation of transcription induced by nuclear GAPDH initiates apoptotic cell death via a network of signaling mechanisms.[185] Once inside the nucleus, there appears to be a change in GAPDH structure associated with oxidative modification of a channel Cys residue (#149 or 150 depending on the species).[187,189] It has been suggested that this modification might be a signal for transcriptional activation of its own gene, but there is no evidence for this at present.

GAPDH binds to unusual oligopeptides that are found in neurodegenerative diseases, but the relevance of this is unclear. Polyglutamine-repeat regions localized in cell nuclei correlate to disease progression and severity in several neurological conditions. Some proteins are known to selectively bind $(Gln)_n$ strands, and one of those is GAPDH. Consequently, even though the neurological effects of GAPDH/$(Gln)_n$ accumulation in cell nuclei are currently unknown, there is an open possibility that this may have causative deleterious effects. Similarly, in Alzheimer's disease, GAPDH binds the

cytoplasmic carboxyl terminus of the β-amyloid protein,[191] and the significance of this is also unresolved.

Overall, there are many possibilities for ways in which GAPDH could be perturbed in therapeutic approaches, particularly in view of the unknowns surrounding its role in the onset and progression of neurological diseases. Our hypothesis is that the quaternary structure of GAPDH may influence the role of this protein in programmed apoptosis, impacting accumulation of the enzyme in the nucleus and what it does there. We are intrigued by the observation that the dimeric forms do not induce apoptotic activities, even though they are *more* active in glycolysis because this supports our supposition that interface mimics to perturb the dimerization state of GAPDH may selectively effects apoptosis in neurodegeneration (*cf*, when CGP3466 binds rabbit GAPDH *in vitro*, it converts the tetramer to a dimeric form,[192] and that is *more* active than the parent tetramer in glycolysis).[193] The fact that CGP3466 gives 100x the apoptotic rescuing effect of deprenyl may be because it changes the enzyme to the dimeric form *more* effectively via a different allosteric binding mode. In other words, perhaps deprenyl has less influence on the interface region than CGP3466, converts it to the dimeric form less effectively, and gives less of an apoptotic rescuing effect.[192] Our preliminary studies have uncovered an opportunity to prepare small molecule interface mimics to perturb assembly and persistence of GAPDH monomers into dimers-of-dimers, and we propose to test compounds that are designed to disrupt these interface regions.

Data From EKO and eKO

There are two types of interfaces in the GAPDH tetramer; one composed of mainly sheet regions that appears to be the most important for binding, and another where the interface is a much less well-defined loop. Overlay of template **1** on the loop region was only observed when the RMSD constraint was relaxed to approximately 0.7; we judged this fit to be unsatisfactory, so that interface is not considered further. However, the fit of template **1**, and the stereoisomers of this, on the other interface gave some *excellent* matches.

Mining of the database in the usual way showed overlay (RMSD 0.28 Å) of the core molecule **1** (LLL) on an interface region of the *human* GAPDH (pdb: 3gpd in which NAD^+ is also bound).[194] Specifically, it overlays in a parallel fashion (*i.e. N*-termini of protein and peptidomimetic are head-to-head) with residues Phe232-Met230-Thr228. These are located on a strand of a β-sheet at the hydrophobic interface formed with a β-sheet on the other monomer (Figure S7a). Consideration of matches for the same peptidomimetic stereoisomer **1** (LLL) for overlay on the same crystal structure but at slightly higher RMS deviations gave a second hit (Figure S7b). This second hit was interesting in three respects: (i) it overlaid with a different interface region (though also a sheet); (ii) the side-chains involved were on discontiguous residues (K308, 65 residues displaced from D243, and V241); and, (iii) unlike the first hit, the peptidomimetic is *antiparallel* with the primary sequence.

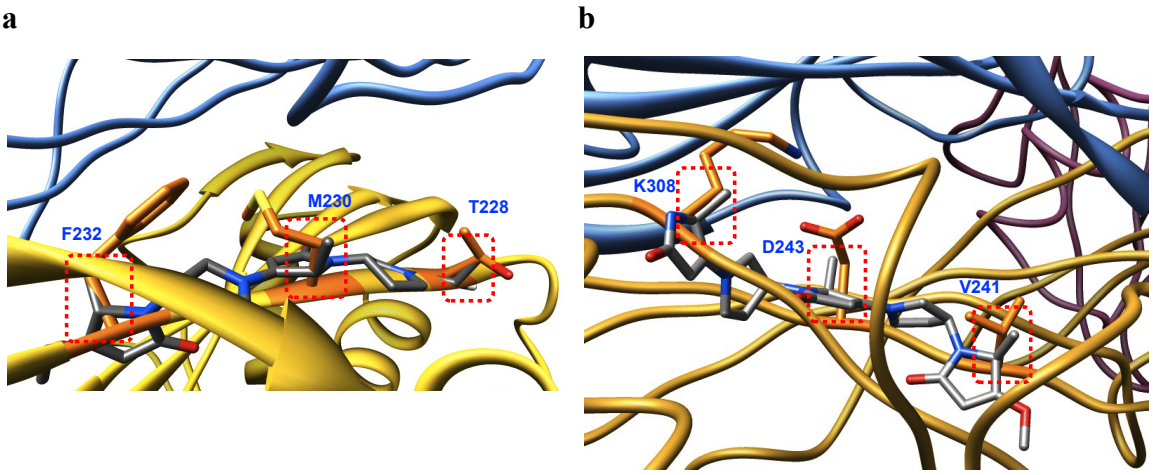


Figure S7. (a) Original; and, (b) secondary hits for overlay of **1** (LLL) on human GAPDH.

Some of the 78 GAPDH structures in the PDB are symmetrical homotetramers, but others have NAD^+ bound to some units.[184] Comparisons (DALI)[195] suggest that the overall fold of GAPDH structures is highly conserved in different organisms.[184]

In the next phase of the mining exercise we modeled all the other stereoisomers (D,L,L-, L,D,L-, L,L,D-, D,D,L-, D,L,D-, L,D,D-, and D,D,D-) of **1** on all the available GAPDH

structures, and many leads emerged (Table S3). Molecules **1** overlaid *N -> C* or *C -> N* with the protein. Interestingly, **1** (LDL) gave more matches than any other isomer, and these could be aligned or inverse oriented. Before synthesis of the compounds, we will check the sequence correspondence between organisms for each of the three amino acid combinations outlined in Table S3, and match them to the source of GAPDH used in the assay (human is preferred; GAPDH is commercially available for all the organisms listed below, rabbit is one of the least expensive, and human is the most).

SAR will be performed around the hits from this first group of compounds. However, all proposed small molecule structures would be matched on GAPDH first to prioritize the syntheses (assuming there are no other factors to be addressed more urgently, e.g. water solubility issues).

Table S3. Interface Mimics Identified using EKO for Overlay on GAPDH

	conformer	PDB	RMSD (Å)	score	residues	directionality	source
1	LLL	3gpd	0.28	12.30	F232-M230-T228	*N -> C*	*human*
2	LLL	1j0x	0.32	16.80	F230-M228-T226	*N -> C*	*rabbit*
3	DLL	1gd1	0.30	18.69	D242-V244-E246	*N -> C*	*bacillus stearothermophilus*
4	LDL	2hki	0.24	14.41	K309-I311-W313	*N -> C*	*spinach*
5	LDL	1znq	0.26	13.40	L177-T179-V181	*N -> C*	*human*
6	LDL	1cer	0.25	12.55	F306-K304-M302	*C -> N*	*thermus aquaticus*
7	LDL	1dc3	0.27	12.54	L171-T173-V175	*N -> C*	*E-coli*
8	LDL	1qxs	0.27	13.19	L189-T191-I193	*N -> C*	*trypanosoma cruzi*
9	LDL	1ml3	0.29	12.54	L189-T191-I193	*N -> C*	*trypanosoma cruzi*
10	LDL	2b4t	0.29	12.55	L183-T185-V187	*N -> C*	*plasmodium falciparum*
11	LDL	1i33	0.29	13.38	L189-T191-I193	*N -> C*	*leishmania mexicana*
12	LDL	1rm3	0.29	13.98	W313-I311-K309	*C -> N*	*spinach*
13	LDL	2prk	0.3	13.39	W313-I311-K309	*C -> N*	*engyodontium album*
14	LDL	1j0x	0.3	17.71	W310-I308-K306	*C -> N*	*rabbit*
15	LLD	1nqa	0.12	7.74	M231-T175-M173	*C -> N*	*bacillus stearothermophilus*
16	LLD	2dbv	0.14	8.09	M231-T175-M173	*C -> N*	*bacillus stearothermophilus*
17	DDL	1cf2	0.29	12.85	T215-V213-I183	*N -> C*	*methanothermus fervidus*

D. Caspases 1 and 3 as Targets for Pyrrolinone-pyrrolidine Oligomers 1

Structure and Function of Caspases 1 and 3

Caspases (**c**ysteinyl **a**spartate-**s**pecific **p**rote**ases**) are intracellular cysteine proteases that specifically cleave substrates at aspartic acid residues. Intracellular modulation of caspases is achieved, in the first instance, by activator (e.g. APAF-1, Fas/FADD) and inhibitor IAPs) proteins. At a second level, the activators are controlled by Bcl-2-family proteins, and the SMAC proteins modulate the inhibitors. Above that level are Bcl-2-family modulators like Bim, Bad, and Bid).[196,197]

Eight of the eleven caspases that are encoded by the human genome are function in apoptosis: these include caspase 3. Two processes turn on caspases: (i) *extrinsic* pathways spurred by activation of cell surface death receptors and mediated by activation of a caspase zymogen by an "up-stream" caspase (e.g. caspase 8); or, (ii) *intrinsic* pathways originating in the mitochondria for which caspase 9 is a typical upstream activator. Signals from both the intrinsic and extrinsic pathways converge at downstream caspases like 3.[198]

Three human caspases activate a subset of proinflammatory cytokines, and these include caspase 1 (of "interleukin-1β-converting enzyme, ICE"). Selective inhibition of caspase 1 prevents production of IL-1β at sites of inflammation. *Activation* of caspase 1, on the other hand, causes mature IL-1β to bind to its type 1 receptor and this plays an important role in mediating neuronal cell death.[199]

Selective inhibition of caspase 1 or 3 could have a range of biochemical consequences. These are difficult to predict, but the evidence that enhanced caspase 1 and 3 activity is associated with many neurological conditions (reviewed several times)[200-202] is indisputable. Ischemic or traumatic injury causes upregulation of caspase 1 and 3 and this has been associated with cell death and neurological deterioration.[203] In Huntington disease, the protein huntingtin is cleaved by caspase 1 and 3 to afford toxic fragments required for the formation of neural intranuclear inclusions and progression of the disease. Inhibition of caspase 1 slows progression of Huntington disease in a mouse

model.[204] Caspase 1 is also implicated in stroke, ALS, and Parkinson's disease.[205] In Alzheimers, caspase 3 (and others) cleave the β-amyloid precursor protein (β-APP) giving a *C*-terminal fragment that is found in senile plaques and is a potent inducer of apoptosis.[205]

For these reasons there has been a great deal of interest in small molecule inhibitors of caspases.[197,205-207] Nonpeptide inhibitors of caspase 3 inhibit apoptosis and maintain cell functionality.

Caspases 1 and 3 have a "dimer of heterodimers" quaternary structure wherein each heterodimer consists of a p10 and p20 fragment. One active site is formed on each p10/p20 interface, hence there are two in the overall quaternary structure of each molecule. Active site His^{237} and Cys^{285} residues are located on the larger, p20, fragments, and the substrate-binding cavity is completed by the protein-protein interface between this and the p10 fragment. Consequently, caspase 1 is a dimer of obligatory heterodimers, wherein dissociation of the p10 and p20 fragments will negate the activity of the enzyme.[208] A natural inhibitor of caspase 1, the serpin "crmA" acts by opening the p10/p20 interface, and possibly that between p10 and p20 too.[209] Consequently, we propose small molecule disruptors of PPIs in caspases 1 and 3 can be used to modulate their activities.

Caspases 1 and 3 have a "dimer of heterodimers" quaternary structure wherein each heterodimer consists of a p10 and p20 fragment. One active site is formed on each p10/p20 interface, hence there are two in the overall quaternary structure of each molecule. Active site His^{237} and Cys^{285} residues are located on the larger, p20, fragments, and the substrate-binding cavity is completed by the protein-protein interface between this and the p10 fragment. Consequently, caspases 1 and 3 are a dimer of obligatory heterodimers, wherein dissociation of the p10 and p20 fragments will negate the activity of the enzyme.[208] A natural inhibitor of caspase 1, the serpin "crmA" acts by opening the p10/p20 interface, and possibly that between p10 and p20 too.[209]

Wells and co-workers used their tethering strategy to identify an allosteric site that is found in caspases.[210,211] This is about 15 Å from the substrate-binding cavity, and it

impacts both the p10/p10 and the p10/p20 interfaces without causing dissociation. Binding of that inhibitor traps the enzyme in an inactive conformation. In *human* caspase 1 that allosteric site is Arg^{286}, Cys^{331} (used for tethering), and Glu^{390}.[50] This is particularly relevant here because the compounds that we find may disrupt interfaces in caspase 1.

Data From EKO and eKO

Caspase 1

Several small molecules were found that impact the p10/p20 interface. In fact, *all* the 8 stereomers of **1** considered gave good overlays (RMSD < 0.22 – 0.46 Å) with slightly different, but overlapping, amino acid tracts in this region. The fact that several parts of this sheet are involved enhances the scope for small molecule interface mimic design. For instance, using the **1** (LDL) framework, residues Asn337-Ser339-Arg341 overlaid on a loop region of the p10 units (*N* -> *C*; one, shown in Figure S8b, was identified in 4 different structures: 1rwk, 1rwn, 1rww, 1bmq).[210] This region is near the active site of the p20 unit, but not directly impacting it. Two matches at slightly higher RMSD were also found (Figure S8c and d); these overlaid with overlapping sheet regions of the p10 unit, Thr388-Phe330-Ile328 (*C*-to-*N*) and Ile328-Phe330-Ser332 (*N*-to-*C*) that interacts with a sheet on the p20 fragment (Figure S8c and d). One of these matches alternates residues (Thr388-Phe330-Ile328) bridging two strands, whereas the other (Ile328-Phe330-Ser332) is aligned with only one. This type of variance validates our concept of "universal mimic" design.

All the stereomers of **1** also gave good overlays for the p10/p10 interface (RMSD 0.27 – 0.42 Å). Thus, **1** (LDL) overlaid with Glu390-Thr388-Met386 (*C-to-N*; central region of Figure 8a, and in e) close to the glutamate residue identified by Wells as binding their allosteric inhibitor.

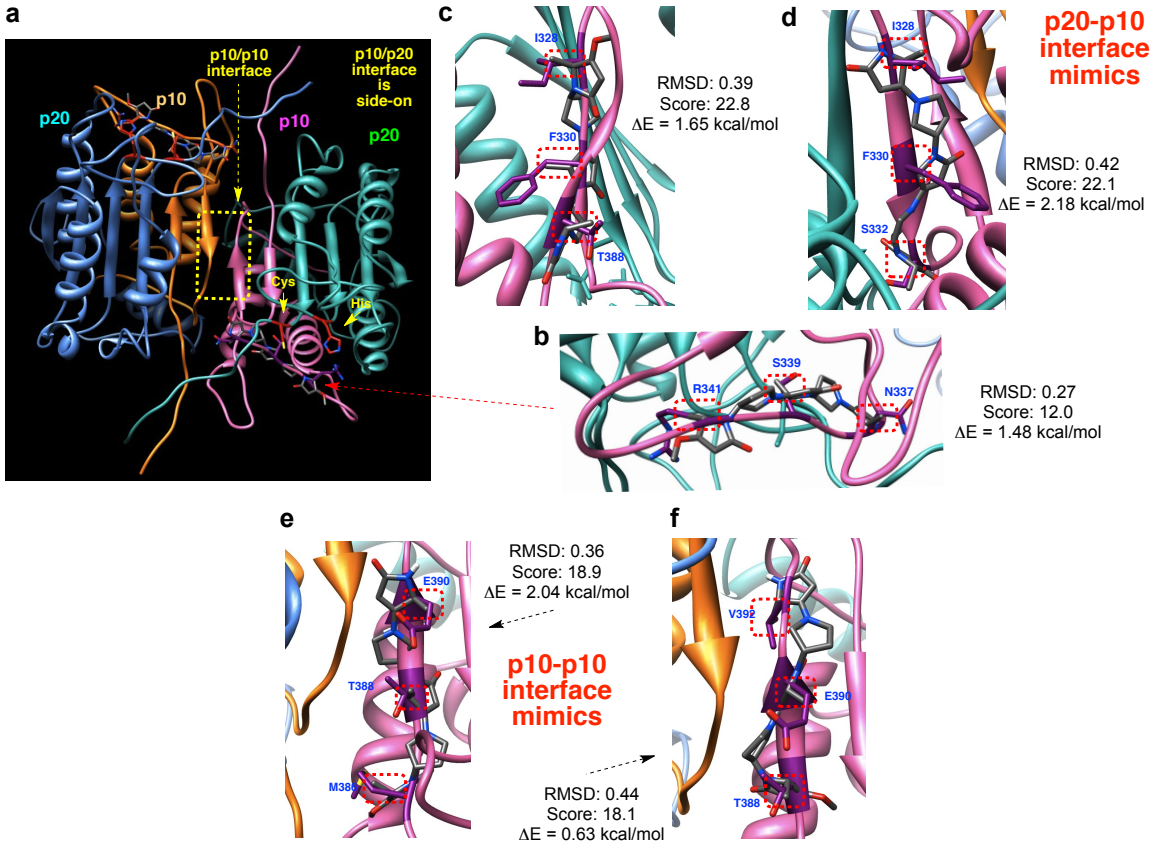


Figure S8. (a) structure of caspase **1** showing p10/p10 and p10/p20 interface; (b) – (d) overlays of **1** (LDL) on the p10/p20 interface; and, (e) – (f) **1** (LDL) on the p10/p10 interface.

Table S4. Interface Mimics Identified using EKO for Overlay on Caspase 1.

	conformer	RMSD (Å)	score	residues	direction	chains interface
1	LLL	0.46	25.5	N337-S339-R341	*N* -> *C*	p20-p10
2	DLL	0.35	15.3	I328-F330-S332	*C* -> *N*	p20-p10
3	LDL	0.22	12.1	N337-S339-R341	*N* -> *C*	p20-p10
4	LLD	0.33	21.8	D326-I328-T389	*N* -> *C*	p20-p10
5	LDD	0.38	23.5	E390-I328-D326	*C* -> *N*	p20-p10
6	DLD	0.29	13.1	I328-F330-S332	*N* -> *C*	p20-p10
7	DDL	0.28	19.1	R391-A329-C331	*C* -> *N* -> *C*	p20-p10
8	DDD	0.32	17.3	E390-I328-D326	*C* -> *N*	p20-p10
9	LLL	0.28	16.8	N337-T334-M386	*C* -> *N* -> *C*	p10-p10
10	DLL	0.41	17.3	T395-K325-I323	*C* -> *N*	p10-p10
11	LDL	0.36	18.9	E390-T388-M386	*C* -> *N*	p10-p10
12	LLD	0.35	21.4	E378-M386-T334	*N* -> *C* -> *N*	p10-p10

13	LDD	0.36	21.1	T334-M386-F377	*N -> C ->N*	p10-p10
14	DLD	0.33	20.3	P380-E378-R376	*C -> N*	p10-p10
15	DDL	0.27	15.5	P380-E378-R376	*C -> N*	p10-p10
16	DDD	0.42	19.8	R376-T388-E390	*N -> C*	p10-p10

Caspase 3

EKO indicated a good overlay of **1** (DDL) on the Ala227-Tyr274-Glu272 sequence in a *N-to-C-to-N* orientation at the p17/p12 interface.[212] Again, remarkably, *all* the stereoisomers of **1** gave good overlays with caspase 3 at this interface (RMSD 0.29 – 0.42 Å). Similarly, matches could be found for the p12/p12 interface but the RMSD values were higher (0.33 – 0.50) indicating slightly inferior correspondence (not shown). Interface mimics corresponding to this region will be prepared, but priority will be given to p17/p12 (which is more interesting anyway because it forms the active site).

Table S5. Interface Mimics Identified using EKO for Overlay on Caspase 3.

	conformer	RMSD (Å)	score	residues	direction	chains interface
1	LLL	0.39	18.9	W206-I216-S213	*N -> C ->N*	p17-p12
2	DLL	0.37	17.2	Y195-F193-E190	*C -> N*	p17-p12
3	LDL	0.37	15.4	R207-F215-L219	*N -> C*	p17-p12
4	LLD	0.42	23.1	L271-L269-L194	*C -> N*	p17-p12
5	LDD	0.31	10.2	C264-A196-L194	*C -> N*	p17-p12
6	DLD	0.30	19.2	F193-Y195-C264	*N -> C*	p17-p12
7	DDL	0.29	12.0	A227-Y274-E272	*N -> C ->N*	p17-p12
8	DDD	0.33	14.8	A196-L194-L269	*C -> N ->C*	p17-p12
9	LLL	0.35	20.7	Y203-I262-A200	*N -> C ->N*	p12-p12
10	DLL	0.39	19.8	A244-K260-A258	*N -> C ->N*	p12-p12
11	LDL	0.39	20.9	K260-I262-N240	*N -> C ->N*	p12-p12
12	LLD	0.50	26.8	K260-I262-C264	*N -> C*	p12-p12
13	LDD	0.42	23.4	V266-M268-T270	*N -> C*	p12-p12
14	DLD	0.33	15.3	C264-V266-M268	*N -> C*	p12-p12
15	DDL	0.45	21.1	C264-I262-K260	*C -> N*	p12-p12
16	DDD	0.39	15.1	N240- I262-K260	*N -> C ->N*	p12-p12

E. α-Antithrombin as Targets for Pyrrolinone-pyrrolidine Oligomers 1

Structure and Function of α-Antithrombin

Serpins are a fascinating group of serine protease inhibitors that are active in their monomeric forms, but can revert to *inactive* fibril-like oligomers. Formation of these fibrils is an undesirable characteristic associated with a series of diseases collectively known as "serpinopathies".[213] Serpinopathies are driven by conformational changes to proteins that lead to fibrils, in ways that parallel, but are different to, amyloid formation in Alzheimer's disease. Overall, interaction of one serpin unit with another, a PPI, governs these events.

A key feature of serpin oligomerizations is that the monomeric proteins are metastable;[214] they revert to thermodynamically more favorable (*ca* 32 kcal/mol)[215] dimeric, then oligomeric forms via a domain swapping process. This involves opening of the proteins via release of a loop region that is intimately associated with a β-strand arrangement in the monomeric form. There is apparently a significant kinetic barrier to formation of the dimeric form, but once this is reached, it opens a gateway to oligomerization. Thus, dimer formation in serpinopathies has been described to impart "infectivity".[216-218] Discovery of a small molecule that can modulate such processes for one serpin would have ramifications for all serpinopathies.[214]

Intriguingly, like Alzheimer's, several serpinopathies are associated with neurological diseases. These include, for instance, involvement of neuroserpin in the formation of "Collins bodies", a characteristic of familial encephalopathy. However, probably the most studied of the serpins is α-antitrypsin; mutated forms (the "*Z*-mutant") of this protein are associated with liver cirrohosis and emphysema. Of particular interest here is another serpin called antithrombin. Mutations of antithrombin are associated with thrombosis, and blood-clotting events in thrombosis are related to stroke.

Various groups have investigated how *peptides* corresponding to the loop region involved in domain swapping can be used to inhibit oligomer formation in serpinopathies. For instance, this approach has been proven for α-antitrypsin,[216-220] and

α-antithrombin.[221] *In vitro* assays used to identify the active peptides in these studies involve differentiation between serpins in monomeric and oligomeric forms. This can be done by gel-electrophoresis and by methods that rely on intrinsic Trp fluorescence.

Ultimately, peptides that are active *in vitro* are unlikely to be useful *in vivo* due to the usual reasons associated with bioavailability (cellular and stability to proteases; oral is not required since intravenous injection of therapeutics for life threatening disease is standard and acceptable). Consequently, the awaits for *small molecules* to be discovered that exert similar inhibition of dimerization properties. This is analogous to the stage of development of Alzheimer's therapies when peptide leads were shown to inhibit amyloid formation.[222]

Data From EKO and eKO

Domain swapping processes leading to antithrombin oligomers involves a loop-sheet interaction in the monomeric closed form being transformed into a similar one between one or more protein monomers. Data mining experiments for this study were performed using the only available structural information (human antithrombin, 2znh), and that involves the *wild type* antithromin and not the mutated one that most inclined to form firbils.[215] Hit PPI regions where EKO predicts compounds **1** could bind (see below) correspond to the sheet region where that the loop interacts with in the closed form. Antithrombin mutations that lead to fibril formation are *not* associated with this loop-sheet interaction and, because of this, the inhibitors that are designed here should be appropriate for mutated serpin.

Table S6 summarizes the interface mimics **1** found after data mining all the stereoisomers; all but one overlaid on the sheet region that is either side of the key loop (Figure S9); the exception (Table S6, entry 5) overlaid on an ill-defined helix-loop motif. Figure S9b shows how **1** (LLL) (small letters correspond to codes for parent amino acids with the same side-chains) overlays with the *C*-terminus of the mimic on the *C*-terminal end of the featured interface; in other words, the two chains that are overlaid run in the *same* direction, so we call this an *N -> C* mimic (natural orientation). Overlays for the

other mimics listed in Table S6 are *superior* to this in terms of RMSD; the worst match is shown here (Figure S9b) but, even so, the correspondence looks fine.

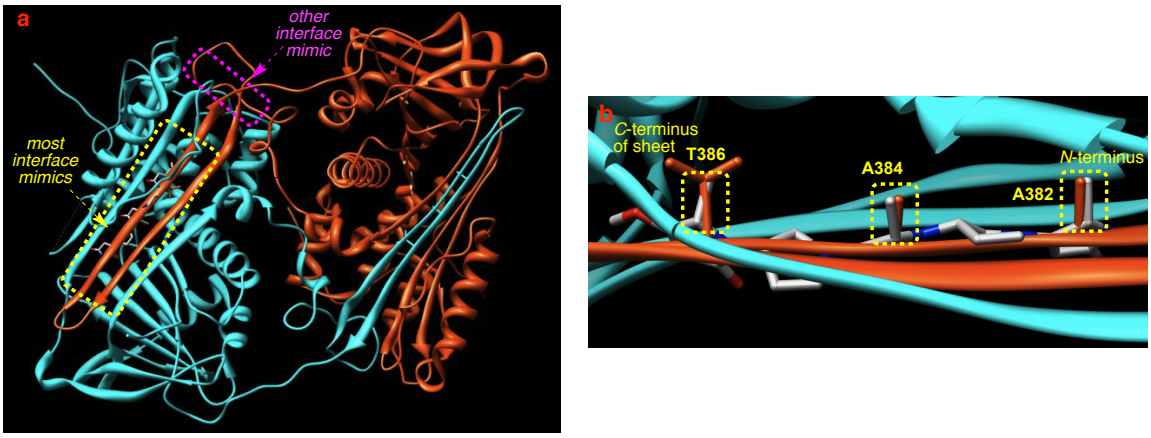


Figure S9. (a) Crystal structure of antithrombin wild-type stable dimer that illustrates how chain swapping can lead to oligomerization and fibril formation, deactivating the serpin; (b) overlay of **1** (LLL) on Ala382-Ala384-Thr386 wherein the sheet and interface mimic orient site the same *N* -> *C* polarity.

Table S6 shows that some interface mimics **1** may "align" the protein strand and others "oppose" it (*N* -> *C*, and *C* -> *N* respectively) but both can give good side-chain fitting. One compound, **1** (DLL) (entry 2; primed small letters indicate D-stereochemistry) overlays with *discontiguous* amino acids that reverse relative to the mimic (*C* -> *N* -> *C*).

Table S6. Interface Mimics Identified using EKO for Overlay on Antithrombin.

	conformer	RMSD (Å)	score	residues	direction
1	LLL	0.37	15.2	A382-A384-T386	*N* -> *C*
2	DLL	0.28	17.6	E374-F372-A384	*C* -> *N* -> *C*
3	LDL	0.25	15.1	S385-H369-A367	*C* -> *N*
4	LLD	0.34	21.8	L373-A383-S385	*N* -> *C*
5	DDL	0.36	19.5	D97-C95-A20	*C* -> *N*
6	DLD	0.33	17.0	V388-T386-K370	*C* -> *N*
7	LDD	0.23	12.6	A384- A382- E374	*C* -> *N*
8	DDD	0.34	14.5	H369-A387-V389	*C* -> *N*

VITA

Name: Eunhwa Ko

Permanent Address: Department of Chemistry
2001 Texas A&M University
P.O. Box 30012
College Station, TX 77842-3012

Education:

2012 Ph.D. Chemistry
Texas A&M University
College Station, TX 77843
Advisor: Dr. Kevin Burgess

2001 MSc. Chemistry
Seoul National University
Gwanak-ro, Gwanak-gu
Seoul 151-742
South Korea
Advisor: Dr. Junghun Suh

2001 BSc. Chemistry
Chungnam National University
99 Daehak-ro, Yuseong-gu
Daejeon 305-764
South Korea